

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τεχνικές Βαθείας Μάθησης για Πολυφασματική Ανάλυση Δορυφορικών Εικόνων



Ηλίας Τσίχλης

Εξεταστική Επιτροπή

Αναπληρωτής Καθηγητής Μιχαήλ Γ. Λαγουδάκης (Σχολή ΗΜΜΥ)

Καθηγητής Μιχάλης Ζερβάκης (Σχολή ΗΜΜΥ)

Αναπληρωτής Καθηγητής Παναγιώτης Παρτσινέβελος (Σχολή ΜΗΧΟΠ)

Χανιά, Απρίλιος 2021

TECHNICAL UNIVERSITY OF CRETE, GREECE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Deep Learning Techniques for Multispectral Satellite Image Analysis



Ilias Tsichlis

Thesis Committee

Associate Professor Michail G. Lagoudakis (School of ECE)

Professor Michalis Zervakis (School of ECE)

Associate Professor Panagiotis Partsinevelos (School of MRE)

Chania, April 2021

Περίληψη

Στον τομέα της Μηχανικής Μάθησης, η ανάγκη κατανόησης της λειτουργίας των βαθιών νευρωνικών δικτύων γίνεται όλο και πιο σημαντική. Είναι γεγονός ότι η προσοχή των ερευνητών επικεντρώνεται κυρίως στο σύνολο δεδομένων που χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσης νευρωνικών δικτύων, καθώς και στην ποιότητα της εξόδου που παράγουν. Ωστόσο, αυτή η εστίαση δεν έχει χρησιμοποιηθεί ειδικά για την καλύτερη κατανόηση της εσωτερικής λειτουργίας τους. Η έρευνα στην παρούσα διπλωματική εργασία αποσκοπεί στον προσδιορισμό του τρόπου με τον οποίο αυτή η εστίαση μπορεί να βελτιώσει τη γνώση του τρόπου λειτουργίας των βαθιών νευρωνικών δικτύων. Συγκεκριμένα, δίνεται έμφαση στις αρχιτεκτονικές δύο γνωστών και ευρέως χρησιμοποιούμενων νευρωνικών δικτύων, VGG16 και ResNet50. Σε αυτό το πλαίσιο, για να εξετασθεί η υπόθεση ότι τα νευρωνικά δίκτυα μπορούν να γίνουν κατανοητά μακροσκοπικά, η παρούσα εργασία εξετάζει τη δημιουργία ενός κατάλληλου, δυναμικού συνόλου δεδομένων για την εκπαίδευση των δύο δικτύων, αλλά και τον προσδιορισμό των βασικών συσχετισμών που καθορίζουν την απόφαση ταξινόμησης. Συγκεκριμένα, έχει αναπτυχθεί ένα εργαλείο προγραμματισμού που έχει τη δυνατότητα να εξάγει δυναμικά δορυφορικές εικόνες από τη βάση δεδομένων OpenStreetMap για να σχηματίσει σύνολα εκπαίδευσης για το πρόβλημα της σωστής ταξινόμησης εικόνων που περιέχουν στοιχεία ανανεώσιμων πηγών ενέργειας (φωτοβολταϊκά, ανεμογεννήτριες, φράγματα). Επιπλέον, χρησιμοποιήθηκαν θερμικοί χάρτες για να εξηγήσουν τη λειτουργία των νευρωνικών δικτύων που εκπαιδεύτηκαν με αυτά τα σύνολα δεδομένων σε συνδυασμό με μια νέα μετρική απόδοσης, η οποία φαίνεται να είναι κατάλληλη για το σκοπό της βαθύτερης διερεύνησης των ταξινομήσεων νευρωνικών δικτύων. Ένα βασικό εύρημα από αυτήν την εργασία είναι ότι, κατά τη διάρκεια της εκπαίδευσης και της ταξινόμησης, αναπτύσσονται ορισμένοι δόκιμοι ή αδόκιμοι συσχετισμοί μεταξύ των χαρακτηριστικών εισόδου και στόχου. Δίνουμε έμφαση στο κατά πόσον αυτές οι πληροφορίες μπορούν να φανούν χρήσιμες, κατά την επιλογή ενός κατάλληλου ταξινομητή για την επίλυση συγκεκριμένων προβλημάτων. Τα αποτελέσματα αυτής της εργασίας δείχνουν ότι η έρευνα, η εξέλιξη και η επιλογή μιας βαθιάς αρχιτεκτονικής νευρωνικών δικτύων για την επίλυση σύγχρονων προβλημάτων ταξινόμησης είναι επιλογές που μπορούν να καθοδηγηθούν αξιοποιώντας πληροφορίες που εξάγονται σύμφωνα με το προαναφερόμενο βασικό εύρημα.

Abstract

In the field of Machine Learning, the need to understand the operation of deep neural networks is becoming more and more important. It is well established that researchers' attention focuses mainly on the dataset used during neural network training, as well as on the quality of the output they produce. However, this focus has not been utilized specifically for better understanding of their internal operation. The research in this diploma thesis aims to determine how this focus can improve the knowledge of how deep neural networks operate. Specifically, emphasis is placed on the architectures of two well-known and widely-used neural networks, VGG16 and ResNet50. In this context, to test the hypothesis that neural networks can be understood macroscopically, this thesis examines the creation of an appropriate, dynamic dataset for training the two networks, but also the identification of the key correlations that determine the classification decision. Specifically, a programming tool has been developed that has the ability to dynamically export satellite images from the OpenStreetMap database to form training sets for the problem of correctly classifying images containing elements of renewable energy sources (photovoltaics, wind turbines, dams). Additionally, heatmaps were used to explain the operation of the neural networks trained with these datasets in conjunction with a novel performance metric, which seems to be appropriate for the purpose of a deeper investigation into neural network classifications. A key finding from this work is that, during training and classification, certain proper or improper correlations between input and target characteristics are being developed. Our emphasis is placed on whether this information can be useful, for choosing an appropriate classifier to solve specific problems. The results of this work indicate that research, evolution, and selection of a deep neural network architecture to solve modern classification problems are choices that can be guided using information extracted according to the mentioned key finding.

Acknowledgements

Throughout the writing of this thesis I have received great support and help from the four people who follow.

First of all, I would like to thank my advisor, Associate Professor Michail G. Lagoudakis, for the opportunity he gave me, as well as for the encouraging way he advised me to do this work.

I would also like to thank Associate Professor Panagiotis Partsinevelos, because not only from the very first moment he accepted me to work with him, but also in all stages of this thesis his advices, the moral support he offered, as well as his patience combined with his wonderful character were invaluable to me.

Last, but not least, I would like to deeply thank two of my very good friends, Trigakis Dimitrios and Chatziparaschis Dimitrios, who from the beginning, with the knowledge they possess in the field, as well as the generosity that expresses them, guided me until the end.

Περιεχόμενα

1	Εισαγωγή	1
1.1	Αντικείμενο της Διπλωματικής Εργασίας	2
1.2	Συνεισφορά της Διπλωματικής Εργασίας	2
1.3	Επισκόπηση της Διπλωματικής Εργασίας	3
2	Ανασκόπηση Θεωρίας	5
2.1	Βασικές έννοιες των Νευρωνικών Δικτύων	5
2.1.1	Δυαδική Ταξινόμηση και Λογιστική Παλινδρόμηση	5
2.1.2	Απώλεια και Συνάρτηση Κόστους	6
2.1.3	Αλγόριθμος Καθόδου Κλίσης	7
2.1.4	Νευρωνικό Δίκτυο Πολλών Επιπέδων	8
2.1.5	Συναρτήσεις Ενεργοποίησης	9
2.1.6	Αλγόριθμος Οπισθοδρόμησης Σφάλματος	11
2.1.7	Φαινόμενα Εξαφάνισης και Έκρηξης των Κλίσεων	12
2.1.8	Βελτιστοποιητές	13
2.1.9	Αρχικοποίηση Παραμέτρων	15
2.1.10	Κανονικοποίηση Παραμέτρων Εισόδου	15
2.2	Συνελικτικά Νευρωνικά Δίκτυα	16
2.2.1	Τεχνικές Συμπλήρωσης Εικόνας	16
2.2.2	Καθορισμός του Βήματος των Συνελίξεων	17
2.2.3	Τεχνικές και Επίπεδα Υποδειγματοληψίας	18
2.2.4	Συνέλιξη Εικόνων Πολλών Διαστάσεων	20
2.2.5	Εφαρμογή Πολλαπλών Φίλτρων	21
2.2.6	Απλή Αρχιτεκτονική Νευρωνικού Δικτύου	21
2.2.7	Αρχιτεκτονικές Νευρωνικών Δικτύων Ευρείας Χρήσης	23

ΠΕΡΙΕΧΟΜΕΝΑ

2.2.8	Αύξηση Δεδομένων Εκπαίδευσης	26
2.2.9	Μεταφορά της Μάθησης Νευρωνικών Δικτύων	27
2.2.10	Χάρτες Χαρακτηριστικών Συνελικτικού Νευρωνικού Δικτύου	28
2.2.11	Χάρτες Ενεργοποίησης των Κλάσεων	30
2.3	Η Μετρική F-score	32
3	Διατύπωση του Προβλήματος	35
3.1	Ανάγκη Κατανόησης των Νευρωνικών Δικτύων	35
3.2	Μέθοδοι Διεύρυνσης του Συνόλου Δεδομένων	36
3.3	Σύνδεση Νευρωνικών Δικτύων και ΑΠΕ	36
3.4	Σχετικές Εργασίες	37
3.4.1	Μέθοδοι Απεικόνισης των Χαρακτηριστικών	37
3.4.2	Μέθοδοι Επιβεβαίωσης των Προβλέψεων	40
4	Η Προσέγγισή μας	43
4.1	Ανάλυση Συστήματος Συντεταγμένων	43
4.2	Η Εφαρμογή Maps Static	45
4.3	Overpass API, QL και OSM Tile Calculator	46
4.4	Εξαγωγή Προσαρμοσμένων Δεδομένων	48
4.5	Δημιουργία Μετρικής Fmatch	53
4.6	Εκπαίδευση με Προσαρμοσμένα Σύνολα Δεδομένων	56
5	Αποτελέσματα	61
5.1	Απόδοση των Αρχιτεκτονικών VGG16, ResNet50	61
5.1.1	Απόδοση του VGG16 στην Ταξινόμηση ΑΠΕ.	62
5.1.2	Απόδοση του ResNet50 στην Ταξινόμηση ΑΠΕ.	65
5.2	Επεξήγηση Δικτύων με Χρήση Θερμικών Χαρτών	68
6	Συμπεράσματα	75
6.1	Συμπεράσματα	75
6.2	Μελλοντική Δουλειά	76
6.2.1	Διερεύνηση Περισσότερων Αρχιτεκτονικών	76
6.2.2	Συνδυάζοντας Όλα τα Στάδια	76
	Βιβλιογραφία	79

Κατάλογος σχημάτων

2.1	Στο διάγραμμα παρουσιάζονται οι τιμές της συνάρτησης κόστους J σε σχέση με τα W και b . Με πράσινα βέλη σχεδιάζονται τα βήματα που ακολουθεί ο αλγόριθμος καθόδου κλίσης.	7
2.2	Διάγραμμα λογιστικής παλινδρόμησης.	8
2.3	Βαθύ νευρωνικό δίκτυο K επιπέδων.	8
2.4	Νευρωνικά δίκτυα ως προς το βάθος.	8
2.5	Σιγμοειδής συνάρτηση.	9
2.6	Υπερβολική Εφαπτομένη.	9
2.7	Συνάρτηση ReLu.	10
2.8	Συνάρτηση Leaky ReLu.	10
2.9	Εξαρτήσεις κατά τους υπολογισμούς του αλγορίθμου Οπισθοδρόμησης Σφάλματος.	12
2.10	Βήματα κατά την διαδικασία της κανονικοποίησης.	16
2.11	Παράδειγμα όμοιας συνέλιξης.	17
2.12	Παράδειγμα βήματος ($s = 4$).	18
2.13	Μέγιστη συγκέντρωση.	19
2.14	Μέση συγκέντρωση.	19
2.15	Συνέλιξη πολλών διαστάσεων ($s=2$)	20
2.16	Περισσότερα φίλτρα για την εξαγωγή πολλών χαρακτηριστικών	21
2.17	Απλή Αρχιτεκτονική Νευρωνικού Δικτύου.	23
2.18	Αρχιτεκτονική του VGG16.[1]	24
2.19	Μη υπολειπόμενο μπλοκ.	25
2.20	Υπολειπόμενο μπλοκ.	26
2.21	Ανεμογεννήτρια από προσαρμοσμένο σύνολο δεδομένων.	28

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

2.22	Δεύτερο επίπεδο.	29
2.23	Τέταρτο επίπεδο.	29
2.24	Δέκατο επίπεδο.	29
2.25	Δέκατο τρίτο επίπεδο.	29
2.26	Χάρτες χαρακτηριστικών από τα εσωτερικά επίπεδα του δικτύου.	29
2.27	Υδροηλεκτρική ενέργεια.	30
2.28	Αιολική ενέργεια.	30
2.29	Θερμικοί χάρτες από προσαρμοσμένο σύνολο δεδομένων.	30
2.30	Ολική μέση συγκέντρωση.	31
3.1	Χάρτες Χαρακτηριστικών του δευτέρου επιπέδου των δύο δικτύων που χρησιμοποιήθηκαν. Η εικόνα προέρχεται από την [2].	37
3.2	Επίπεδο του deconvnet το οποίο είναι ενωμένο παράλληλα με τον αντίστοιχο επίπεδο του νευρωνικού δικτύου. Με αυτό το επίπεδο ανακατασκευάζονται τα χαρακτηριστικά του προηγούμενου επιπέδου. Η εικόνα προέρχεται από την [3].	38
3.3	Μια αναπαράσταση της αντίθετης συγκέντρωσης. Ο προσδιορισμός της τοποθεσίας των μεγίστων γίνεται με τους διακόπτες. Η εικόνα προέρχεται από την [3].	39
3.4	Γραφική Αναπαράσταση της μεθόδου Guided Grad-CAM. Η εικόνα προέρχεται από την [4].	40
3.5	Παραδείγματα της μεθόδου Guided Grad CAM. Η εικόνα προέρχεται από την [4].	41
3.6	Μεθόδος LIME. Η εικόνα προέρχεται από την [5].	41
3.7	Απεικόνιση των τριών μεγαλύτερων προβλέψεων στη μέθοδο LIME. Η εικόνα προέρχεται από την [5].	42
4.1	Παγκόσμιες συντεταγμένες.	44
4.2	Αναπαράσταση συντεταγμένων των πλακιδίων του χάρτη.	45
4.3	Χρέωση για την εφαρμογή Static Map.	46
4.4	Η εφαρμογή OSM Tile Calculator.	47
4.5	Εξαγωγέας Δεδομένων.	50
4.6	Επεξεργασία δεδομένων κατά το πρώτο επίπεδο.	50
4.7	Πλακάκι με ένα ground_truth_bbox.	51

4.8	Συνδυασμός δεδομένων και λήψη δορυφορικών εικόνων με χρήση των Mid-Coordinates.	51
4.9	Δεδομένα εξόδου.	52
4.10	Δορυφορικές εικόνες από το σύνολο δεδομένων ΑΠΕ.	52
4.11	Στις δύο πρώτες εικόνες που ο συσχετισμός έγινε με βάση την σκίαση, η μετρική έχει χαμηλότερη τιμή σε σύγκριση με τις επόμενες δύο.	53
4.12	Στις δύο πρώτες εικόνες που ο συσχετισμός έγινε βασισμένος στο συνδυασμό νερού και δάσους, η μετρική έχει χαμηλότερη τιμή σε σύγκριση με τις δύο πρώτες.	54
4.13	VGG16 με προσθήκη επιπέδου ολικής μέσης συγκέντρωσης για το σύνολο δεδομένων γεννητριών ΑΠΕ.	57
4.14	Υπερπαράμετροι για το VGG16.	57
4.15	Προσαρμογή και αξιολόγηση του VGG16.	58
4.16	ResNet50 με προσθήκη επιπέδου ολικής μέσης συγκέντρωσης για το σύνολο δεδομένων γεννητριών ΑΠΕ.	59
5.1	Απώλεια Cross Entropy - Μετρική Fbeta.	62
5.2	Πίνακες σύγκρισης για το VGG16.	63
5.3	Ιστογράμματα Μετρικής για το δίκτυο VGG16.	64
5.4	Απώλεια Cross Entropy - Μετρική Fbeta	65
5.5	Πίνακες σύγκρισης για το Resnet50.	66
5.6	Ιστογράμματα Μετρικής για το δίκτυο ResNet50.	67
5.7	Υδροηλεκτρική γεννήτρια.	68
5.8	Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης hydro - VGG16.	69
5.9	Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης hydro - ResNet50.	70
5.10	Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης wind - VGG16.	71
5.11	Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης wind - ResNet50.	72
5.12	Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης solar - VGG16.	73
5.13	Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης solar - ResNet50.	74

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Κεφάλαιο 1

Εισαγωγή

Η Μηχανική Μάθηση και πιο συγκεκριμένα η Βαθιά Μάθηση, ήδη άλλαξε τον τρόπο με τον οποίο αντιμετωπίζουμε πολλά προγραμματιστικά προβλήματα. Αν συλλογιστούμε πως υλοποιούνται πλέον οι διαφημίσεις, οι αγορές, καθώς και οι αναλύσεις και αναζητήσεις πληροφοριών, θα δούμε πως είναι κατά κόρον τέκνα εφαρμογών Βαθιάς Μάθησης. Ακόμη, και σε θέματα δημιουργίας νέων τεχνολογιών, θέματα υγείας, τέχνης, καθώς και εφαρμογών αυτοματοποίησης η Βαθιά Μάθηση έρχεται πάλι στο προσκήνιο.

Είναι λοιπόν βέβαιο πως στο κοντινό μέλλον η Μηχανική Μάθηση υπόσχεται αναμφίβολα να καταφέρει μεγάλες αλλαγές, ακόμα και στα πιο συνηθισμένα ζητήματα της ζωής ενός ατόμου. Έτσι οι νέοι ερευνητές δεν έχουν μόνο την ευκαιρία, αλλά και την υποχρέωση να συνεισφέρουν στον στόχο αυτό.

Καθίσταται λοιπόν ως απαραίτητη προϋπόθεση η βαθύτατη εξερεύνηση της συμπεριφοράς των νευρωνικών δικτύων, καθώς και των τμημάτων που τα αποτελούν, με απώτερο σκοπό την χάραξη του δρόμου βελτίωσής τους.

1.1 Αντικείμενο της Διπλωματικής Εργασίας

Στην σημερινή εποχή, όλο και περισσότερες είναι εκείνες οι εφαρμογές οι οποίες στηρίζονται σε βαθιά νευρωνικά δίκτυα, επηρεάζοντας τόσο την καθημερινότητα των ανθρώπων όσο και την ίδια την έρευνα νέων τεχνολογιών. Για τον λόγο αυτό σημαντική καθίσταται η ανάγκη βελτίωσης της λειτουργίας τους.

Αν και τα τελευταία χρόνια, η κατανόηση της συμπεριφοράς τους βρίσκεται στο επίκεντρο της έρευνας νέων τεχνολογιών [3], [6], από προηγούμενες εργασίες δεν έχει γίνει με όμοιο τρόπο η επεξήγηση της προσοχής στο κατά πόσο ένα νευρωνικό υλοποιεί τις ταξινομήσεις των κλάσεων, κάνοντας τις συσχετίσεις με τα χαρακτηριστικά των αντικειμένων τους καθώς και με εκείνα που αποτελούν το περιβάλλον τους. Ακόμη, δεν έχει δοθεί έμφαση με ίδιο τρόπο, στην διαδικασία με την οποία δημιουργείται δυναμικά το σύνολο δεδομένων στο οποίο εκπαιδεύονται τα νευρωνικά δίκτυα, καθώς και στην σημασία της σύγκρισης στην απόδοση και στην λειτουργία μεταξύ αυτών.

Έτσι, ο συνδυασμός των παραπάνω συνιστά μια νέα μέθοδο, η οποία όχι μόνο διευκολύνει την καλύτερη επιλογή νευρωνικού δικτύου κατά την επίλυση συγκεκριμένου προβλήματος, αλλά ορίζει και την κατεύθυνση για την εξέλιξή του.

1.2 Συνεισφορά της Διπλωματικής Εργασίας

Η συνεισφορά αυτής της εργασίας κατευθύνεται στην σημασία του συνόλου δεδομένων, που χρησιμοποιείται κατά την εκπαίδευση ενός νευρωνικού και κατά πόσο αυτό διαμορφώνει την λειτουργία του, καθώς και στην επεξήγηση του τρόπου με τον οποίο τελικά ένα δίκτυο ταξινομεί. Ακόμη, ορίζεται μια μέθοδος, για τον τρόπο με τον οποίο συγκρίνονται διαφορετικά γνωστά νευρωνικά δίκτυα, για την επίλυση ταξινομήσεων δορυφορικών εικόνων.

Αρχικά, το πρώτο ζήτημα εκ των τριών, επιτυγχάνεται με την υλοποίηση εργαλείου για την εξαγωγή συνόλου δεδομένων δορυφορικών εικόνων με δυναμικό τρόπο. Το σύστημα αυτό χρησιμοποιεί πληροφορίες από την βάση δεδομένων του OpenStreetMap και δημιουργήθηκε έτσι ώστε η μορφή των δεδομένων που εξάγει, αλλά και ο τρόπος από τον οποίο προέρχονται οι δορυφορικές εικόνες, να προσδιορίζονται από τις ανάγκες του χρήστη. Έπειτα, το δεύτερο ζήτημα προσεγγίζεται, με την επεξήγηση θερμικών χαρτών των κλάσεων

που ταξινομεί το μοντέλο και με την δημιουργία κατάλληλης μετρικής για τον προσδιορισμό της ευστοχίας των ταξινομήσεων. Τέλος, η σύγκριση γνωστών νευρωνικών δικτύων γίνεται με τα δίκτυα VGG16 και ResNet50 τα οποία εκπαιδεύτηκαν σε σύνολο δορυφορικών εικόνων, οι οποίες απεικονίζουν γεννήτριες ΑΠΕ.

1.3 Επισκόπηση της Διπλωματικής Εργασίας

Στο Κεφάλαιο 2 παρουσιάζουμε όλες τις πληροφορίες της θεωρίας που ήταν απαραίτητο να μελετηθούν για την υλοποίηση της διπλωματικής εργασίας. Συγκεκριμένα, δίνεται έμφαση στις θεμελιώδεις αρχές των βασικών και των συνελικτικών νευρωνικών δικτύων, των θερμικών χαρτών και της μετρικής F-score. Στο Κεφάλαιο 3 γίνεται η διατύπωση του προβλήματος της διπλωματικής εργασίας, και επίσης παρουσιάζονται σχετικές εργασίες οι οποίες συσχετίζονται με την κατεύθυνση που ακολουθήσαμε. Στο Κεφάλαιο 4 δίνεται η μεθοδολογία που ακολουθήθηκε κατά την διεκπεραίωση αυτής της εργασίας. Στο Κεφάλαιο 5 παρουσιάζουμε την απόδοση των νευρωνικών δικτύων που εκπαιδεύσαμε, την σύγκριση των συναρτήσεων κόστους, την σύγκριση των λειτουργιών μεταξύ αυτών, την επεξήγηση της συμπεριφοράς τους, καθώς και τις τιμές που πετυχαίνουν σε σχέση με τη μετρική που δημιουργήσαμε. Τέλος, το Κεφάλαιο 6 αποτελεί τον επίλογο αυτής της εργασίας, δίνοντας έμφαση στα βασικότερα συμπεράσματα που προέκυψαν και σε μελλοντικούς στόχους που ακολουθούν.

1. ΕΙΣΑΓΩΓΗ

Κεφάλαιο 2

Ανασκόπηση Θεωρίας

2.1 Βασικές έννοιες των Νευρωνικών Δικτύων

2.1.1 Δυαδική Ταξινόμηση και Λογιστική Παλινδρόμηση

Η δυαδική ταξινόμηση καλύπτει το πρόβλημα στο οποίο θέλουμε να δούμε αν ένα στοιχείο ανήκει σε μια κλάση ή όχι. Επομένως, αν το στοιχείο εισόδου ανήκει στην κλάση, τότε έχουμε τιμή εξόδου $y = 1$, ενώ αν δεν ανήκει έχουμε τιμή εξόδου $y = 0$. Η σημειογραφία για την εκμάθηση σε δυαδική ταξινόμηση, με πλήθος m παραδειγμάτων, ορίζεται ως (X, Y) , όπου $X \in R^{n \times m}$ είναι η είσοδος του συστήματος, και $Y \in R^{n \times m}$ είναι οι τιμές που θα συγκριθούν με τις προβλέψεις εξόδου του συστήματος και έχουν τιμές 0 ή 1. Ένας αλγόριθμος για την επίλυση της δυαδικής ταξινόμησης είναι η λογιστική παλινδρόμηση.

Στην πραγματικότητα, η λογιστική παλινδρόμηση είναι ο αλγόριθμος εκμάθησης που χρησιμοποιούμε όταν οι τιμές εξόδου Y ανήκουν σε πρόβλημα με ισχυρό πλαίσιο επιβλεπομένης μάθησης και έχουν τιμές 0 ή 1 (δυαδική ταξινόμηση). Εφόσον έχουμε ένα διάνυσμα χαρακτηριστικών εισόδου X , θέλουμε έναν αλγόριθμο που μπορεί να εξάγει την πρόβλεψη \hat{y} , η οποία είναι η εκτίμηση μας για το y , για να την συγκρίνουμε με την πραγματική. Δηλαδή, θέλουμε την πιθανότητα το y να είναι ίσο με 1 δεδομένων των χαρακτηριστικών μίας εισόδου x ($\hat{y} = P(y = 1|x)$).

Με διαστάσεις μεμονωμένης εισόδου $x \in R^{n \times 1}$, οι παράμετροι του συστήματος είναι $w \in R^{n \times 1}$ και $b \in R$, όπου w τα βάρη και b οι σταθερές. Έτσι, η έξοδος του συστήματος υπολογίζεται από την φόρμουλα $\hat{y} = \sigma(w^T \times x + b)$, όπου σ είναι η σιγμοειδής συνάρτηση. Επομένως, ο στόχος για την επίλυση του προβλήματος είναι να βρούμε παραμέτρους βαρών

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

(w) και σταθερών (b) ώστε οι προβλέψεις εξόδου (\hat{y}) να είναι όσο πιο κοντά στις πραγματικές τιμές (y) των παραδειγμάτων εκμάθησης.

2.1.2 Απώλεια και Συνάρτηση Κόστους

Για την εκπαίδευση των παραμέτρων W και B του μοντέλου της λογιστικής παλινδρόμησης ορίζεται η συνάρτηση απώλειας. Στόχος είναι να μετρηθεί πόσο καλή είναι η πρόβλεψη \hat{y} σε σχέση με την πραγματική τιμή εξόδου y . Η συνάρτηση απώλειας για την λογιστική παλινδρόμηση έχει την μορφή: $L(\hat{y}, y) = -(y \times \log(\hat{y}) + (1 - y) \times \log(1 - \hat{y}))$.

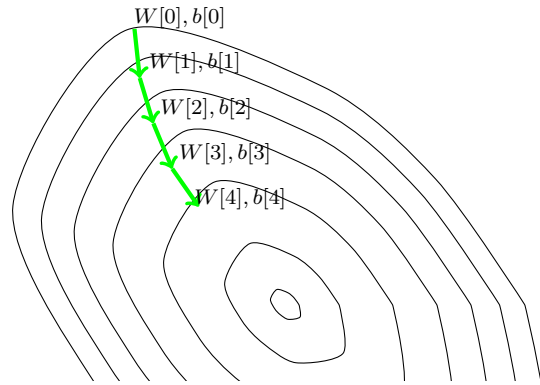
Η συγκεκριμένη συνάρτηση απώλειας χρησιμοποιείται διότι στην περίπτωση που η τιμή του y είναι 1 τότε η συνάρτηση παίρνει την μορφή $L(\hat{y}, y) = -\log(\hat{y})$. Οπότε, για να έχουμε μικρό σφάλμα πρέπει η τιμή της πρόβλεψης να είναι μεγάλη (κοντά στην μονάδα). Από την άλλη, αν η τιμή του y είναι 0, τότε η συνάρτηση παίρνει την μορφή $L(\hat{y}, y) = -\log(1 - \hat{y})$. Έτσι για να έχουμε μικρό σφάλμα, πρέπει η τιμή της πρόβλεψης να είναι μικρή (κοντά στο μηδέν).

Η συνάρτηση απώλειας ορίζεται μόνο για ένα παράδειγμα εκπαίδευσης (μετράει πόσο είναι το σφάλμα όταν έχουμε μόνο μια είσοδο). Για πολλές εισόδους ορίζεται η συνάρτηση κόστους η οποία υπολογίζει την τιμή του σφάλματος για όλα τα παραδείγματα εκπαίδευσης του μοντέλου. Η συνάρτηση κόστους για την λογιστική παλινδρόμηση έχει τη μορφή $J(W, B) = \frac{1}{m} \times \sum_{i=1}^m L(\hat{y}^i, y^i)$, για πλήθος παραδειγμάτων εκπαίδευσης m .

2.1.3 Αλγόριθμος Καθόδου Κλίσης

Χρησιμοποιούμε τον αλγόριθμο καθόδου κλίσης για να εκπαιδεύσουμε τις παραμέτρους W και B του μοντέλου. Γνωρίζουμε ότι η συνάρτηση κόστους μετρά πόσο καλά γίνονται οι προβλέψεις του μοντέλου, βάσει των παραμέτρων W και B . Θέλουμε λοιπόν να βρούμε W και B που κάνουν τη συνάρτηση κόστους $J(W, B)$ να έχει όσο το δυνατόν μικρότερη τιμή.

Στο Διάγραμμα 2.1 παρουσιάζονται οι παράμετροι W και b στον χώρο (στην πράξη το W είναι πολύ μεγαλύτερο διάνυσμα από το b). Η συνάρτηση κόστους δημιουργεί μια επιφάνεια παίρνοντας τις τιμές της ως προς τα W και b . Στόχος είναι να βρεθούν τιμές για W και b που αντιστοιχούν στο ελάχιστο της συνάρτησης κόστους J .



Σε αυτό το σημείο επισημαίνεται πως η συνάρτηση κόστους πρέπει να είναι κυρτή συνάρτηση, επειδή μη κυρτές συναρτήσεις έχουν πολλά διαφορετικά σημεία τοπικών βελτιστοποιήσεων.

Αρχικά, ο αλγόριθμος καθόδου κλίσης ξεκινά από κάποιο σημείο κάνοντας βήματα προς την κοντινότερη κατηφορική κατεύθυνση. Έτσι, επαναλαμβάνοντας την ακόλουθη ενημέρωση των βαρών και των σταθερών του μοντέλου προχωράει με μικρά βήματα μέχρι να επιτευχθεί σύγκλιση της συνάρτησης κόστους.

Εικόνα 2.1: Στο διάγραμμα παρουσιάζονται οι τιμές της συνάρτησης κόστους J σε σχέση με τα W και b . Με πράσινα βέλη σχεδιάζονται τα βήματα που ακολουθεί ο αλγόριθμος καθόδου κλίσης.

- $W := W - \alpha \frac{\partial J(W, b)}{\partial W}$
- $b := b - \alpha \frac{\partial J(W, b)}{\partial b}$

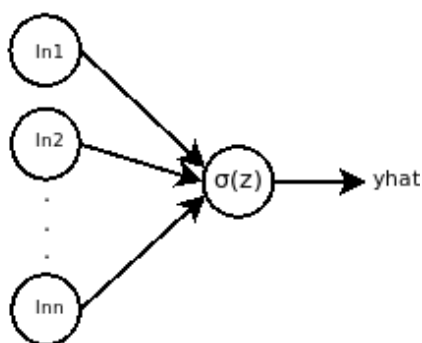
Στις δύο αυτές εξισώσεις, το α αποτελεί τον ρυθμό μάθησης του μοντέλου, ελέγχοντας έτσι πόσο είναι το μέγεθος των βημάτων που κάνει ο αλγόριθμος σε κάθε βήμα.

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

2.1.4 Νευρωνικό Δίκτυο Πολλών Επιπέδων

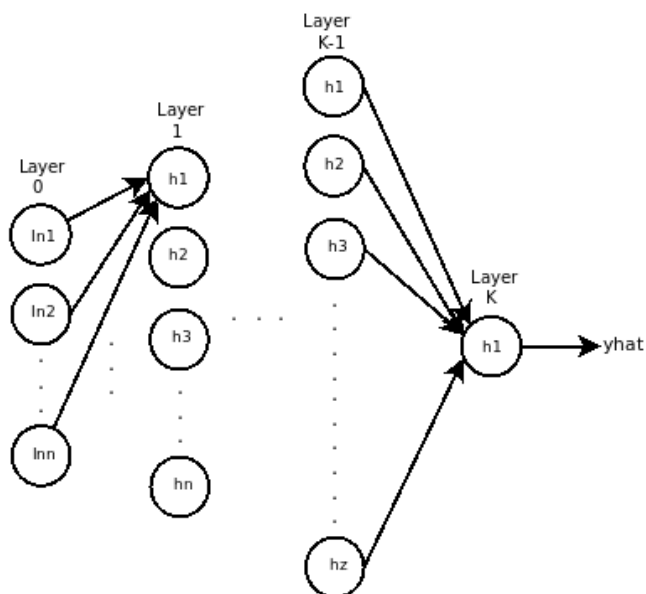
Απλά νευρωνικά δίκτυα, με μικρό αριθμό κρυφών επιπέδων, όπως η λογιστική παλινδρόμηση που έχει μόνο ένα επίπεδο, όπως απεικονίζεται στο Σχήμα 2.2, δεν επιλύουν πολυπλοκότερα προβλήματα ταξινόμησης. Έτσι υπάρχει η ανάγκη για δίκτυα με πολλά περισσότερα επίπεδα. Αυτό συμβαίνει επειδή βαθύτερα δίκτυα μπορούν να μάθουν πολύπλοκες συναρτήσεις για τον διαχωρισμό των δεδομένων στο χώρο. Ένα βαθύ νευρωνικό δίκτυο με K επίπεδα απεικονίζεται στο Σχήμα 2.3.

logistic regression



Εικόνα 2.2: Διάγραμμα λογιστικής παλινδρόμησης.

Deep Neural Network with K layers.



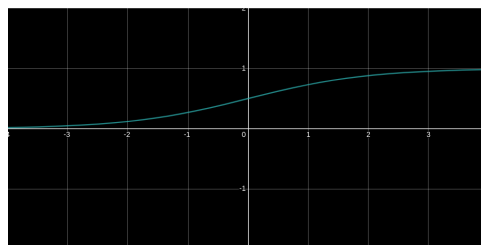
Εικόνα 2.3: Βαθύ νευρωνικό δίκτυο K επιπέδων.

Εικόνα 2.4: Νευρωνικά δίκτυα ως προς το βάθος.

2.1.5 Συναρτήσεις Ενεργοποίησης

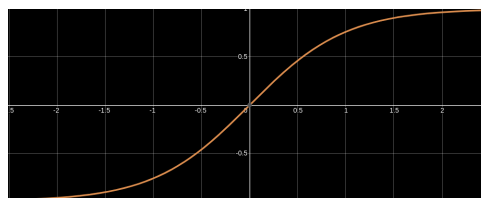
Κατά την σχεδίαση ενός νευρωνικού δικτύου, μία από τις βασικές επιλογές που πρέπει να κάνουμε είναι εκείνη των συναρτήσεων ενεργοποίησης που θα χρησιμοποιηθούν στα κρυμμένα επίπεδα και στις εξόδους του δικτύου.

Αρχικά, μία από τις βασικότερες συναρτήσεις ενεργοποίησης είναι η μη γραμμική σιγμοειδής συνάρτηση 2.5 που έχει τύπο $\sigma(z) = \frac{1}{(1+e^{-z})}$, με σύνολο τιμών στο εύρος $(0, 1)$. Η παράγωγος της είναι $\sigma'(z) = \sigma(z)(1 - \sigma(z)) = \frac{e^{-z}}{(e^{-z}+1)^2}$.



Εικόνα 2.5: Σιγμοειδής συνάρτηση.

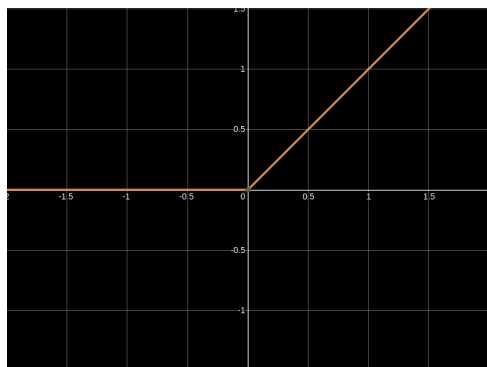
Τις περισσότερες φορές χρησιμοποιούμε διαφορετική συνάρτηση ενεργοποίησης από την σιγμοειδή, διότι σχεδόν πάντα λειτουργούν καλύτερα. Μία από αυτές τις συναρτήσεις ενεργοποίησης είναι η υπερβολική εφαπτομένη 2.6 (tanh) η οποία έχει τύπο $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ και παράγωγο $\tanh'(z) = 1 - \frac{e^z - e^{-z}}{(e^z + e^{-z})^2}$.



Εικόνα 2.6: Υπερβολική Εφαπτομένη.

Η συνάρτηση αυτή έχει σύνολο τιμών $(-1, +1)$ και είναι μια μετατοπισμένη εκδοχή της σιγμοειδούς. Έχει αποδειχθεί, ότι για τις μονάδες των εσωτερικών επιπέδων, η tanh λειτουργεί καλύτερα από την σιγμοειδή. Αυτό γίνεται διότι ο μέσος όρος των τιμών ενεργοποίησης που εξέρχονται από ένα εσωτερικό επίπεδο με συνάρτηση ενεργοποίησης tanh έχει κέντρο το μηδέν. Έτσι κατά την διαδικασία της εκπαίδευσης, αυτό το γεγονός κάνει τους υπολογισμούς της μάθησης ευκολότερους. Η μόνη εξαίρεση που η σιγμοειδής συνάρτηση ενεργοποίησης προτιμάται στην θέση της tanh είναι για το επίπεδο εξόδου όταν θέλουμε η πρόβλεψη μας να είναι μεταξύ 0 και 1 (πρόβλημα δυαδικής ταξινόμησης).

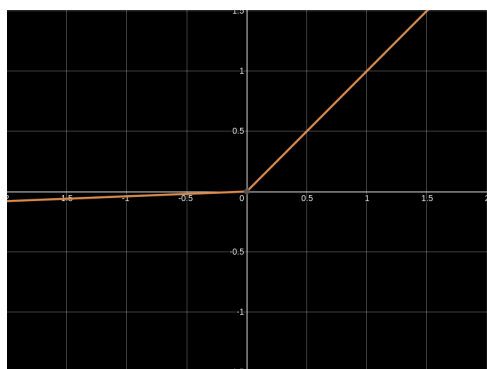
2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ



Εικόνα 2.7: Συνάρτηση ReLu.

Το βασικό μειονέκτημα αυτών των δύο παραπάνω συναρτήσεων είναι ότι αν η μεταβλητή z είναι αρκετά μεγάλη ή μικρή, τότε η παράγωγος (κλίση) αυτών των συναρτήσεων γίνεται σχεδόν μηδέν. Αυτό το γεγονός επιβραδύνει την διαδικασία εκμάθησης. Μια άλλη επιλογή είναι η συνάρτηση ενεργοποίησης ReLu 2.7. Η συνάρτηση ReLu έχει τύπο $a = \max(0, z)$ και η παράγωγος της έχει τιμή 1 για θετικές τιμές της μεταβλητής z , και τιμή 0 για αρνητικές τιμές της μεταβλητής z .

Η Relu είναι μια καλή επιλογή ως συνάντηση ενεργοποίησης στα εσωτερικά επίπεδα του δικτύου, καθώς κάνει το μοντέλο να μαθαίνει γρήγορα. Αυτό γίνεται λόγω της μορφής της παραγώγου, αφού το μοντέλο καταλήγει να ενημερώνει μόνο εκείνα τα βάρη τα οποία έχουν θετικές ενεργοποιήσεις.



Εικόνα 2.8: Συνάρτηση Leaky ReLu.

Ένα μειονέκτημα της ReLu είναι το «φαινόμενο των νεκρώσεων της ReLu». Το φαινόμενο αυτό κάνει τους νευρώνες στα βαθύτερα επίπεδα άχρηστους και ο λόγος πίσω από αυτό βρίσκεται στην ύπαρξη του μηδενός της παραγώγου της ReLu που καθιστά το γινόμενο των μερικών παραγώγων μηδενικό (κανόνας αλυσίδας). Μπορούμε να αποφύγουμε το φαινόμενο αυτό, χρησιμοποιώντας μια άλλη έκδοση της συνάρτησης ReLu, την Leaky ReLu 2.8.

Η συνάρτηση Leaky ReLu αντί να είναι μηδέν όταν το z έχει αρνητικές τιμές, παίρνει μια μικρή κλίση κοντά στον αρνητικό άξονα. Ο τύπος της Leaky Relu είναι $a = \max(kz, z)$, όπου k είναι αρκετά μικρή σταθερά.

Ο λόγος για τον οποίο χρησιμοποιούμε μη γραμμικές συναρτήσεις, σχετίζεται με τις δυνατότητες του δικτύου. Έτσι, για να μπορέσει ένα νευρωνικό δίκτυο να υπολογίσει σύνθετες συναρτήσεις για τον διαχωρισμό των δεδομένων εισόδου στον χώρο, πρέπει οι συναρτήσεις ενεργοποιήσεων που θα επιλεγθούν να είναι μη γραμμικές. Ο λόγος για το ότι συμβαίνει αυτό, είναι επειδή ο συνδυασμός δύο ή περισσότερων γραμμικών συναρτήσεων μας δίνει μια γραμμική συνάρτηση ως έξοδο. Έτσι, εάν δεν χρησιμοποιούμε μη γραμμικές συναρτήσεις ενεργοποιήσεων στο δίκτυο, τότε θα προβλέπει απλά εξόδους με βάση μια γραμμική εκτίμηση.

2.1.6 Αλγόριθμος Οπισθοδρόμησης Σφάλματος

Ο αλγόριθμος οπισθοδρόμησης σφάλματος είναι η διαδικασία κατά την οποία υπολογίζουμε τις κλίσεις του δικτύου. Υποθέτοντας ένα δίκτυο με K επίπεδα, με τον αλγόριθμο αυτόν κινούμαστε προς τα πίσω στο δίκτυο για τον υπολογισμό των κλίσεων.

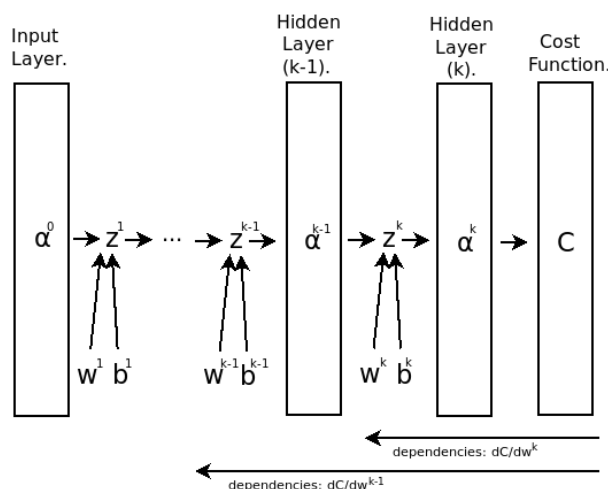
Αρχικά, για το τελευταίο επίπεδο έχουμε δύο εξισώσεις, μία για τα βάρη και μία για τις σταθερές.

- $$\frac{\partial C}{\partial w^k} = \frac{\partial C}{\partial a^k} \frac{\partial a^k}{\partial z^k} \frac{\partial z^k}{\partial w^k}$$
- $$\frac{\partial C}{\partial b^k} = \frac{\partial C}{\partial a^k} \frac{\partial a^k}{\partial z^k} \frac{\partial z^k}{\partial b^k}$$

Αυτές οι εξισώσεις μετρούν την αναλογία με την οποία ένα συγκεκριμένο σύνολο βαρών ή σταθερών (από το τελευταίο επίπεδο) επηρεάζει τη μεταβολή της συνάρτησης κόστους.

Για να υπολογίσουμε τις κλίσεις των προηγούμενων επιπέδων του δικτύου, πρέπει να συνδέσουμε περισσότερες μερικές παραγώγους (κανόνες αλυσίδας). Το σχετικό γράφημα των εξαρτήσεων των συνδέσεων για τα προηγούμενα επίπεδα του δικτύου εμφανίζεται στο Σχήμα 2.9.

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ



Εικόνα 2.9: Εξαρτήσεις κατά τους υπολογισμούς του αλγορίθμου Οπισθοδρόμησης Σφάλματος.

Όπως απεικονίζεται και στο σχήμα, για τον υπολογισμό των κλίσεων σε ένα επίπεδο k_0 , έχουμε εξαρτήσεις μόνο από τα βαθύτερα επίπεδα του νευρωνικού δικτύου. Έτσι, οι γενικοί τύποι για τον υπολογισμό των κλίσεων στο k_0 επίπεδο είναι:

- $\frac{\partial C}{\partial w^{k_0}} = \frac{\partial C}{\partial a^k} \frac{\partial a^k}{\partial z^k} \frac{\partial z^k}{\partial a^{k-1}} \frac{\partial a^{k-1}}{\partial z^{k-1}} \dots \frac{\partial z^{k_0}}{\partial w^{k_0}}$
- $\frac{\partial C}{\partial b^{k_0}} = \frac{\partial C}{\partial a^k} \frac{\partial a^k}{\partial z^k} \frac{\partial z^k}{\partial a^{k-1}} \frac{\partial a^{k-1}}{\partial z^{k-1}} \dots \frac{\partial z^{k_0}}{\partial b^{k_0}}$

2.1.7 Φαινόμενα Εξαφάνισης και Έκρηξης των Κλίσεων

Σε ένα αρκετά βαθύ νευρωνικό δίκτυο στο οποίο χρησιμοποιούμε μια από τις γνωστές συναρτήσεις ενεργοποίησης (όπως η ReLu) όταν τα βάρη γίνουν μεγαλύτερα της μονάδας σε ένα συγκεκριμένο μονοπάτι του δικτύου, τότε οι ενεργοποιήσεις που αντιστοιχούν σε αυτό αυξάνονται εκθετικά σε σχέση με το πλήθος των επιπέδων του δικτύου. Από την αντίθετη πλευρά, αν τα βάρη έχουν τιμές ανάμεσα στο μηδέν και την μονάδα, τότε οι ενεργοποιήσεις μειώνονται εκθετικά.

Το ίδιο μοτίβο ακολουθούν και οι κλίσεις των ενεργοποιήσεων και των βαρών του δικτύου, αφού υπολογίζονται από τα γινόμενα των μερικών παραγώγων που προέρχονται από τα βαθύτερα επίπεδα του δικτύου (κανόνας αλυσίδας). Έτσι προκύπτουν δύο βασικά προβλήματα κατά την εκπαίδευση του δικτύου.

1. Φαινόμενο εξαφάνισης των κλίσεων: Στην περίπτωση που μικραίνουν οι κλίσεις του δικτύου, επειδή οι κλίσεις των πρώτων επιπέδων εξαρτώνται από τις κλίσεις των βαθύτερων, το δίκτυο μαθαίνει με διαφορετικούς ρυθμούς. Για την ακρίβεια, τα πρώτα επίπεδα του δικτύου εκπαιδεύονται πολύ πιο αργά από ότι τα βαθύτερα. Αυτό οδηγεί στην δυσκολία εύρεσης του βέλτιστου σημείου εκπαίδευσης.
2. Φαινόμενο έκρηξης των κλίσεων: Στην περίπτωση που μεγαλώνουν υπερβολικά οι κλίσεις του δικτύου, τότε κατά την εκπαίδευση αλλάζουν υπερβολικά τα βάρη του δικτύου, με αποτέλεσμα να καθίσταται δύσκολη η εύρεση του βέλτιστου σημείου εκπαίδευσης αφού είναι πολύ πιθανό να ξεπερνιέται, καταλήγοντας όλο και σε πιο ακραίες καταστάσεις.

2.1.8 Βελτιστοποιητές

Αλγόριθμος Στοχαστικής κατάβασης των κλίσεων (SGD):

Βασικός αλγόριθμος για την σύγκλιση ενός νευρωνικού δικτύου είναι εκείνος της Στοχαστικής Κατάβασης των κλίσεων. Κατά την λειτουργία του λαμβάνουμε υπόψιν το κόστος που προκύπτει μόνο από ένα τυχαία επιλεγμένο παράδειγμα εκπαίδευσης (μικρό σύνολο δεδομένων) και για τον λόγο αυτόν ανήκει στην κατηγορία των στοχαστικών αλγορίθμων. Αυτή η στοχαστική προσέγγιση συνήσφέρει στην εκμάθηση του δικτύου με αρκετά μικρό κόστος στην ακρίβεια.

Ο αλγόριθμος λειτουργεί με την εφαρμογή των κλίσεων που έχουν υπολογιστεί από την διαδικασία της οπισθοδρόμησης σφάλματος, χρησιμοποιώντας τον τύπο $w = w - \eta \frac{\partial C}{\partial w}$, όπου η είναι ο ρυθμός εκμάθησης και $\frac{\partial C}{\partial w}$ είναι η κλίση για το αντίστοιχο βάρος.

Σε σχέση με νέες προσεγγίσεις, ο αλγόριθμος στοχαστικής κατάβασης των κλίσεων συγκλίνει αρκετά αργά και είναι πιθανό να καταλήξει το νευρωνικό δίκτυο σε τοπικό ελάχιστο των βαρών του σε σχέση με την συνάρτηση κόστους.

Αλγόριθμος Ορμής (Momentum):

Η διαίσθηση για την λειτουργία του αλγόριθμου της Ορμής γίνεται κατανοητή στο ότι για κάθε νέα εποχή (νέος χρόνος εκπαίδευσης) κινούμαστε γρηγορότερα (μεγαλύτερες αλλαγές στις παραμέτρους) από τα προηγούμενα βήματα. Ο τύπος με τον οποίο ο αλγόριθμος

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

ενημερώνει τα βάρη είναι $w_{t+1} = w_t - \eta \frac{\partial C}{\partial w_t} + \gamma \sum_{time=1}^t \eta \frac{\partial C}{\partial w_{time}}$, όπου η μεταβλητή *time* εκφράζει όλες τις περασμένες εποχές και το γ είναι σταθερά περίπου ίση με την μονάδα.

Βλέπουμε πως διατηρεί πληροφορίες για τις αλλαγές των προηγούμενων βημάτων και με τον τρόπο αυτό συγκλίνει ταχύτερα από τον αλγόριθμο στοχαστικής κατάβασης των κλίσεων. Όμως, ακόμα και με τον αλγόριθμο της Ορμής είναι πιθανό να καταλήξουμε σε τοπικό ελάχιστο, διότι στην περίπτωση που αποκτήσουμε μεγάλη ορμή (μεγάλες αλλαγές στις παραμέτρους) ξεπερνάμε το σημείο του ολικού ελαχίστου.

Αλγόριθμος Διάδοσης της Μέσης Τετραγωνικής ρίζας (RMS):

Ο τρόπος λειτουργίας αυτού του αλγορίθμου μοιάζει αρκετά με εκείνη του αλγορίθμου της ορμής, με την κύρια διαφορά να βρίσκεται στην αλλαγή της παραμέτρου μάθησης (η) σε κάθε βήμα του, διαιρώντας την με το $E[g^2]_t$ (εκθετικά φθίνων μέσος όρος).

Η λειτουργία του αλγορίθμου μέσης τετραγωνικής ρίζας βασίζεται στην προσαρμογή του ρυθμού εκμάθησης, περιορίζοντας τις κλίσεις με βάση τα πιο πρόσφατα βήματα. Ο τύπος του είναι $w_{t+1} = w_t - \frac{\partial \eta}{\partial E[g^2]_t} \frac{\partial C}{\partial w}$ όπου το $E[g^2]_t = (1 - \gamma)g^2 + \gamma E[g^2]_{t-1}$, όπου $g = \frac{\partial C}{\partial w_t}$.

Με την προσαρμογή του ρυθμού εκμάθησης λειτουργεί καλύτερα από τον αλγόριθμο της ορμής έχοντας μεγαλύτερη πιθανότητα να βρεθεί σε ολικό ελάχιστο.

Αλγόριθμος Εκτίμησης Προσαρμοσμένης Ροπής (Adam):

Καλός βελτιστοποιητής δεν είναι αυτός που λύνει μόνο ένα μικρό αριθμό προβλημάτων, αλλά είναι εκείνος που γενικεύεται καλά στο ευρύ φάσμα των νευρωνικών δικτύων που ίσως χρειαστεί να εκπαιδευτούν. Ο αλγόριθμος διάδοσης της Μέσης Τετραγωνικής ρίζας είναι μια καλή προσέγγιση για βελτιστοποιητή, όμως έχει αποδειχθεί ότι ο Adam είναι μια καλύτερη επιλογή, αφού λειτουργεί καλά σε ένα ευρύτερο φάσμα αρχιτεκτονικών. Ο Adam επί της ουσίας είναι ο συνδυασμός των αλγορίθμων της ορμής και διάδοσης της μέσης τετραγωνικής ρίζας.

- Όπως και στον αλγόριθμο διάδοσης της Μέσης Τετραγωνικής ρίζας, με την διαφορά ότι έχουμε την υπερπαραμέτρο γ_1 αντί για γ : $S_{dw} = E[g^2]_t = (1 - \gamma_1)g^2 + \gamma_1 E[g^2]_{t-1}$
- Όπως και στον αλγόριθμο της ορμής, με την διαφορά ότι έχουμε την υπερπαραμέτρο γ_2 αντί για γ : $V_{dw} = \gamma_2 \sum_{time=1}^t \eta \frac{\partial C}{\partial w_{time}}$

Επιπροσθέτως, για τον αλγόριθμο Adam έχουμε τις διορθωτικές συναρτήσεις ανά χρονικό βήμα με τύπο $S_{dw} = \frac{S_{dw}}{(1-\gamma 1^t)}$ και $V_{dw} = \frac{V_{dw}}{(1-\gamma 2^t)}$.

Συνδυάζοντας τα παραπάνω, τα βάρη ανανεώνονται ως: $w = w - \eta \times V_{dw} / \sqrt{S_{dw}}$

2.1.9 Αρχικοποίηση Παραμέτρων

Για την εκπαίδευση ενός νευρωνικού δικτύου είναι σημαντικό να αρχικοποιήσουμε τα βάρη τυχαία (ή να τα αρχικοποιήσουμε έτσι ώστε να έχουν διαφορετικές τιμές μεταξύ τους), διότι στην αντίθετη περίπτωση ο αλγόριθμος αναβάθμισης των βαρών δεν παράγει επιθυμητά αποτελέσματα.

Αυτό συμβαίνει επειδή όταν αρχικοποιούνται ως μηδενικά ή με ίδιες τιμές, ο αλγόριθμος αναβάθμισης αλλάζει όλα τα βάρη με τον ίδιο τρόπο. Οπότε σε κάθε βήμα εκπαίδευσης όλοι οι νευρώνες θα υπολογίζουν την ίδια συνάρτηση, πράγμα το οποίο είναι μη επιθυμητό.

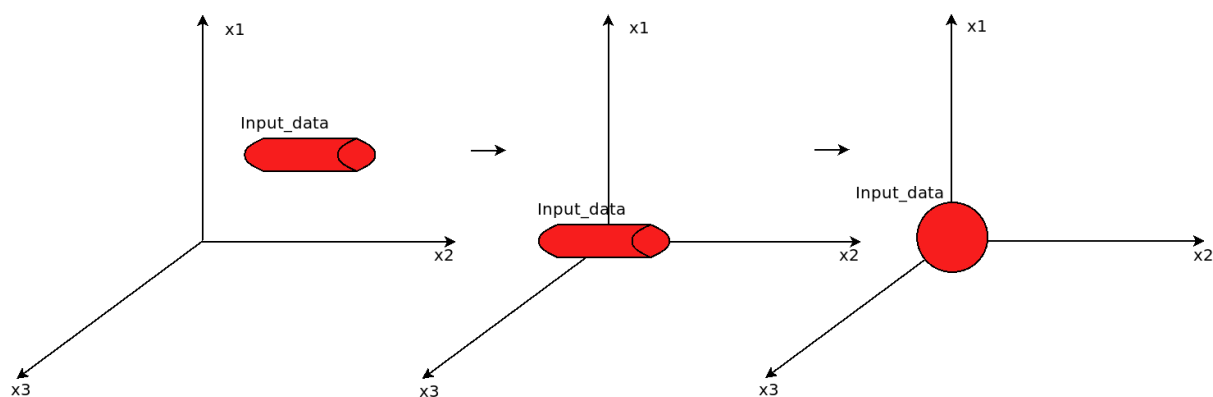
Ακόμη, σχετικά με την ενεργοποίηση των βαρών θέλουμε να αρχικοποιήσουμε τα βάρη με μικρές τιμές, έτσι ώστε να αποφύγουμε τα προβλήματα της εξαφάνισης ή της έκρηξης των κλίσεων (Ενότητα 2.1.7).

2.1.10 Κανονικοποίηση Παραμέτρων Εισόδου

Η κανονικοποίηση των παραμέτρων εισόδου είναι απαραίτητη, αφού στην αντίθετη περίπτωση το σχήμα της συνάρτησης κόστους παίρνει αλλόκοτες μορφές και δεν γνωρίζουμε την τιμή που πρέπει να ορίσουμε για το ρυθμό εκμάθησης. Έτσι, με την κανονικοποίηση των παραμέτρων εισόδου, το σχήμα της συνάρτησης κόστους έχει συμμετρική μορφή και μπορούμε να επιλέξουμε μια μικρή τιμή για τον ρυθμό εκμάθησης.

Η κανονικοποίηση γίνεται σε δύο βήματα, πρώτα μετακινούμε τα δεδομένα, ώστε να έχουν μέση τιμή ίση με μηδέν και στην συνέχεια κανονικοποιούμε ως προς την διακύμανση των δεδομένων εισόδου (Εικόνα 2.10).

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ



Εικόνα 2.10: Βήματα κατά την διαδικασία της κανονικοποίησης.

2.2 Συνελικτικά Νευρωνικά Δίκτυα

2.2.1 Τεχνικές Συμπλήρωσης Εικόνας

Για εικόνα με διαστάσεις (n, n) κάνοντας συνέλιξη με φίλτρο διαστάσεων (f, f) οι διαστάσεις της εξόδου προκύπτουν $(n - f + 1, n - f + 1)$. Τα αρνητικά της παραπάνω διαδικασίας είναι τα ακόλουθα.

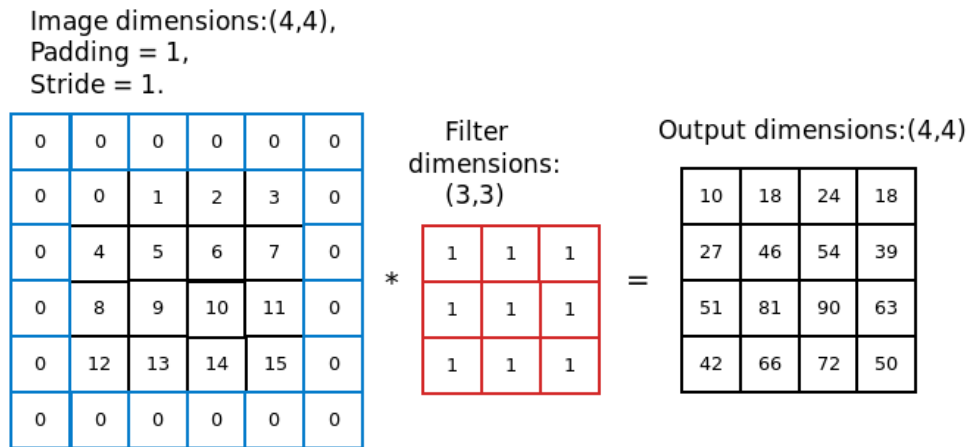
1. Πρώτον, κάθε φορά που εφαρμόζουμε μια συνέλιξη η εικόνα συρρικνώνεται, έτσι μπορούμε να το κάνουμε αυτό λίγες φορές, προτού η εικόνα μας γίνει πολύ μικρή.
2. Δεύτερον, το εικονοστοιχείο στην ακριανή γωνία της εικόνας χρησιμοποιείται μόνο σε ένα στοιχείο της εξόδου. Σε αντίθεση, τα μεσαία εικονοστοιχεία χρησιμοποιούνται σε πολλά. Έτσι, τα εικονοστοιχεία στις άκρες των γωνιών χρησιμοποιούνται πολύ λιγότερο στη δημιουργία της εξόδου, ως αποτέλεσμα μεγάλο μέρος της πληροφορίας να χάνεται κοντά στην άκρη της εικόνας.

Για την επίλυση αυτών των προβλημάτων προτού εφαρμόσουμε την συνελικτική λειτουργία, υλοποιούμε τεχνικές συμπλήρωσης στην εικόνα. Ορίζουμε την παράμετρο p (από το padding) και προχωρούμε στο γέμισμα της εικόνας με ένα επιπλέον περίγραμμα από πλήθος $4 \times p \times n + 4 \times \sum_{i=0}^p (2i + 1)$ εικονοστοιχείων. Με αυτόν τον τρόπο, η έξοδος της συνελικτικής λειτουργίας έχει διαστάσεις $(n + 2p - f + 1, n + 2p - f + 1)$. Ως εκ τούτου, τα εικονοστοιχεία που βρίσκονταν στην γωνία της εικόνας μετατρέπονται σε μεσαία,

επηρεάζοντας πλέον πολλά κελιά της εξόδου. Έτσι, η απώλεια της πληροφορίας στην άκρη της εικόνας μειώνεται. Επιπροσθέτως, με την εφαρμογή αυτής της μεθόδου, μπορούμε να δημιουργήσουμε έξοδο σε οποιοδήποτε μέγεθος χρειαζόμαστε.

Σε σχέση με τις τεχνικές συμπλήρωσης της εικόνας, ορίζονται δύο είδη συνελίξεων, η έγκυρη συνέλιξη και η όμοια συνέλιξη

1. Σε μια έγκυρη συνέλιξη ουσιαστικά δεν χρησιμοποιείται συμπλήρωση, οπότε σε αυτή την περίπτωση αν έχουμε μια εικόνα διαστάσεων (n, n) , κάνοντας συνέλιξη με φίλτρο (f, f) θα έχουμε μια έξοδο με μέγεθος $(n - f + 1, n - f + 1)$.
2. Σε μια όμοια συνέλιξη [2.11](#), το μέγεθος της συμπλήρωσης ορίζεται έτσι ώστε το μέγεθος της εξόδου να είναι ίδιο με εκείνο της εικόνας εισόδου. Οπότε λύνοντας την σχέση $n + 2p - f + 1 = n$ ως προς την μεταβλητή p προκύπτει, $p = (f - 1)/2$.



Εικόνα 2.11: Παράδειγμα όμοιας συνέλιξης.

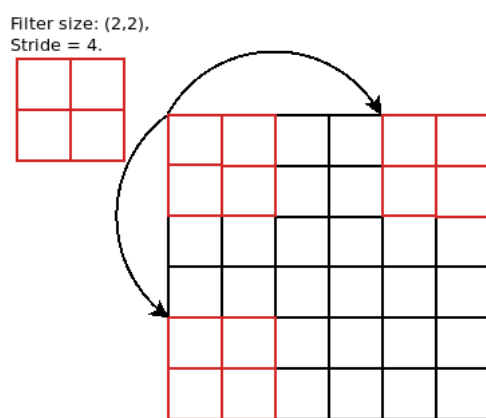
2.2.2 Καθορισμός του Βήματος των Συνελίξεων

Αυτή η τεχνική αποτελεί βασικό κομμάτι της δομής των συνελίξεων, όπως χρησιμοποιούνται στα Συνελικτικά Νευρωνικά Δίκτυα. Ορίζοντας την τιμή του βήματος ως k , υποδηλώνει πως αφού υπολογίσουμε το γινόμενο μεταξύ των βαρών του φίλτρου και των εικονοστοιχείων της επάνω αριστερής γωνίας της εικόνας καθώς και την πρόσθεση μεταξύ αυτών, για τον υπολογισμό του πρώτου εικονοστοιχείου εξόδου, στη συνέχεια αντί να παραβλέψουμε ένα

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

εικονοστοιχείο, θα προχωρήσουμε k βήματα την φορά. Η διαδικασία ακολουθείται και ως προς τις γραμμές και ως προς τις στήλες κατά τις οποίες μετακινείται το φίλτρο στην συνέλιξη όπως φαίνεται στο Σχήμα 2.12.

Εάν έχουμε μια εικόνα με διαστάσεις (n, n) και φίλτρο με διαστάσεις (f, f) , ορίζοντας τιμή συμπλήρωσης ως p και βήματος ως s , τότε οι διαστάσεις της εξόδου είναι $\frac{n - f + 2p}{s} + 1$.



Εικόνα 2.12: Παράδειγμα βήματος ($s = 4$).

2.2.3 Τεχνικές και Επίπεδα Υποδειγματοληψίας

Εκτός από συνελικτικά επίπεδα, τα συνελικτικά νευρωνικά δίκτυα συχνά χρησιμοποιούν και συγκεντρωτικά επίπεδα, με στόχο να μειώσουν το μέγεθος της αναπαράστασης, να επιταχύνουν τον υπολογισμό των εξόδων, καθώς και να καταστήσουν κάποια από τα χαρακτηριστικά που ανιχνεύουν πιο ισχυρά.

Τα δύο πιο συχνά χρησιμοποιούμενα είδη συγκεντρωτικών επιπέδων είναι η μέγιστη και η μέση συγκέντρωση.

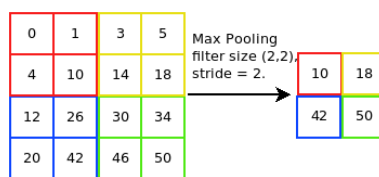
1. Μέγιστη συγκέντρωση 2.13: κάθε στοιχείο της εξόδου είναι το μέγιστο της υποπεριοχής που εφαρμόζεται το φίλτρο δειγματοληψίας.
 - Η διαίσθηση πίσω από αυτό που κάνει η μέγιστη συγκέντρωση σε ότι αφορά τις ενεργοποιήσεις σε κάποιο επίπεδο του νευρωνικού δικτύου, φαίνεται από το γεγονός πως μια μεγάλη τιμή εικονοστοιχείου δηλώνει την ανίχνευση του

συγκεκριμένου χαρακτηριστικού. Οπότε, παίρνοντας τις μέγιστες τιμές των διαχωρισμένων υποπεριοχών της εισόδου, ανιχνεύεται το χαρακτηριστικό στις υποπεριοχές της, όπου διατηρεί έναν υψηλό αριθμό. Όμως, στην περίπτωση όπου το συγκεκριμένο χαρακτηριστικό δεν ανιχνεύεται σε μια υποπεριοχή, τότε η μέγιστη τιμή αυτής της περιοχής είναι ένας μικρός αριθμός.

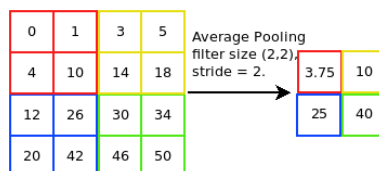
- Οι υπερπαραμέτροι για την μέγιστη συγκέντρωση είναι το μέγεθος φίλτρου (f) και το βήμα δειγματοληψίας (s).

2. Μέση συγκέντρωση 2.14: κάθε στοιχείο της εξόδου είναι ο μέσος όρος της υποπεριοχής που εφαρμόζεται το φίλτρο δειγματοληψίας.

- Σε αντίθεση με την μέγιστη συγκέντρωση, η μέση συγκέντρωση διατηρεί μέση πληροφορία για την ύπαρξη χαρακτηριστικού σε μια υποπεριοχή της εικόνας εισόδου. Παρόλο που η μέγιστη συγκέντρωση έχει περισσότερες εφαρμογές στα νευρωνικά δίκτυα, είναι αρκετές εκείνες οι φορές στις οποίες χρειάζεται να μεταφέρουμε πληροφορία από όλα τα σημεία της εισόδου σε επόμενο επίπεδο του δικτύου. Έτσι, η μέση συγκέντρωση καθίσταται εξίσου σημαντική.
- Οι υπερπαραμέτροι για την μέγιστη συγκέντρωση είναι το μέγεθος φίλτρου (f), καθώς και το βήμα δειγματοληψίας (s).



Εικόνα 2.13: Μέγιστη συγκέντρωση.



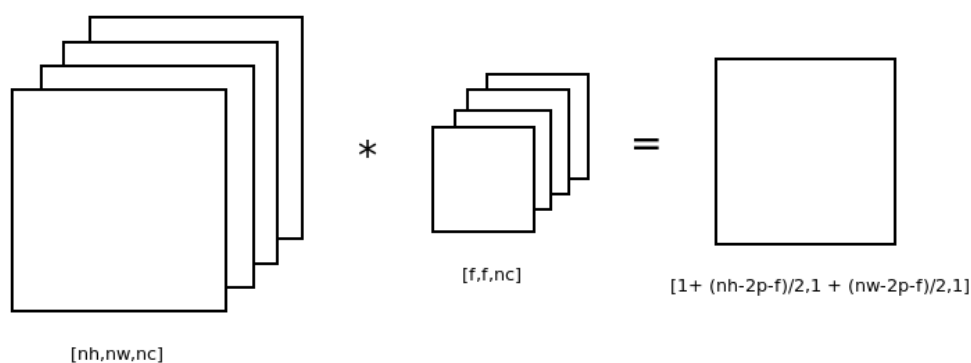
Εικόνα 2.14: Μέση συγκέντρωση.

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

Για διαστάσεις συγκεντρωτικού επιπέδου $[n_h, n_w, n_c]$ όπου n_h η διάσταση ύψους, n_w η διάσταση πλάτους και n_c η διάσταση των καναλιών της εικόνας, τότε οι διαστάσεις της εξόδου υπολογίζονται από τον τύπο: $(1+(n_h-f-2p)/s, 1+(n_w-f-2p)/s, n_c)$. Ο αριθμός των καναλιών εξόδου είναι ίσος με αυτών της εισόδου, διότι η διαδικασία της συγκέντρωσης εφαρμόζεται σε κάθε ένα κανάλι της εισόδου ανεξάρτητα. Τα συγκεντρωτικά επίπεδα δεν έχουν παραμέτρους για μάθηση (βάρος), αλλά μόνο υπερπαραμέτρους.

2.2.4 Συνέλιξη Εικόνων Πολλών Διαστάσεων

Για να ανιχνεύσουμε χαρακτηριστικά όχι μόνο σε μια γκρι μονοδιάστατη εικόνα, αλλά και σε μια πολύχρωμη πολλών διαστάσεων, όπου το πλήθος των καναλιών (n_c) είναι μεγαλύτερο της μονάδας, εφαρμόζουμε την διαδικασία της συνέλιξης με φίλτρο το οποίο έχει πλήθος n_c καναλιών, που αντιστοιχούν στα κανάλια της εικόνας Σχήμα 2.15. Έτσι, ο αριθμός των καναλιών του φίλτρου πρέπει να ταιριάζει με τον αριθμό των καναλιών της εικόνας εισόδου.

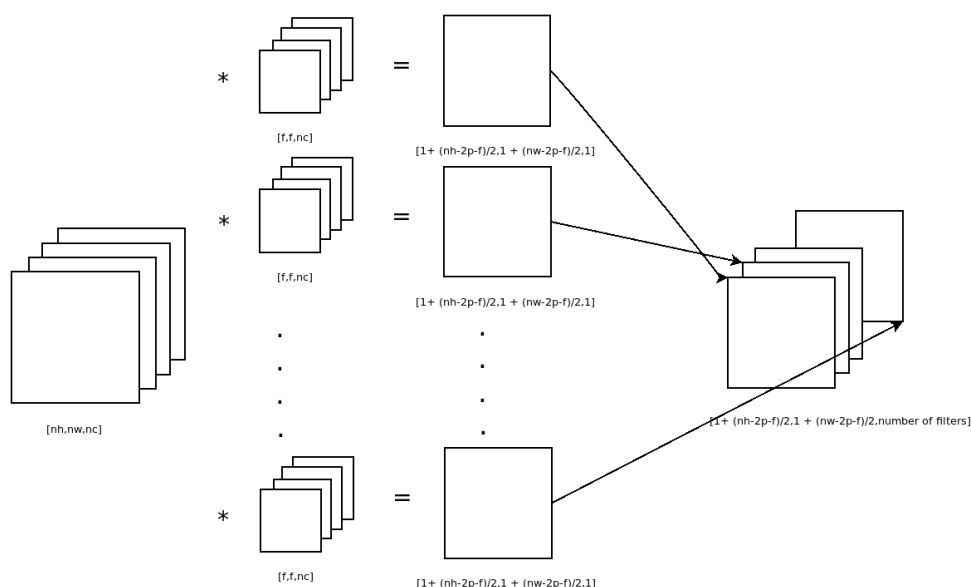


Εικόνα 2.15: Συνέλιξη πολλών διαστάσεων ($s=2$)

Για τον υπολογισμό της εξόδου αυτής της συνέλιξης τοποθετείται φίλτρο διαστάσεων $[f, f, n_c]$ στην επάνω αριστερή γωνία της εικόνας και πολλαπλασιάζεται κάθε στοιχείο του με τα αντίστοιχα εικονοστοιχεία αυτής. Έπειτα τα γινόμενα που προέκυψαν αθροίζονται υπολογίζοντας έτσι το πρώτο κελί της εξόδου. Για το επόμενο κελί, ολισθαίνουμε το φίλτρο κατά s βήματα και επαναλαμβάνεται η διαδικασία.

2.2.5 Εφαρμογή Πολλαπλών Φίλτρων

Κύρια προϋπόθεση για την κατασκευή συνελικτικών νευρωνικών δικτύων είναι η εφαρμογή πολλαπλών φίλτρων. Σε περίπτωση που δεν θέλουμε απλώς να ανιχνεύσουμε μόνο ένα χαρακτηριστικό, αλλά επιζητούμε την ανίχνευση πολλών διαφορετικών χαρακτηριστικών, χρησιμοποιούμε πολλαπλά φίλτρα ταυτόχρονα.



Εικόνα 2.16: Περισσότερα φίλτρα για την εξαγωγή πολλών χαρακτηριστικών

Οπότε, όπως απεικονίζεται στο Σχήμα 2.16, εφαρμόζοντας k φίλτρα στην εικόνα εισόδου δημιουργείται έξοδος με διάσταση βάθους (k) ίση με το πλήθος των φίλτρων που εφαρμόστηκαν.

2.2.6 Απλή Αρχιτεκτονική Νευρωνικού Δικτύου

Έχοντας εικόνα με διαστάσεις ύψους και πλάτους $n_h = n_w = 34$ και πλήθος καναλιών n_c , μπορεί να γίνει αναγνώριση αν φτιαχτεί το ακόλουθο συνελικτικό δίκτυο εμπνευσμένο από το δίκτυο LeNet-5 [9].

Ορίζοντας διαστάσεις των πρώτων φίλτρων $(7, 7)$, βήμα δειγματοληψίας ίσο με μονάδα ($s = 1$) και χωρίς συμπλήρωση εικόνων εισόδου, οι διαστάσεις των ενεργοποιήσεων στο επίπεδο αυτό χρησιμοποιώντας πλήθος 8 φίλτρων προκύπτουν $(28, 28, 8)$, όπως έχει ήδη

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

εξηγηθεί. Ονομάζοντας το επίπεδο αυτό *conv1* και προσθέτοντας τώρα τις σταθερές, καθώς και εφαρμόζοντας μια συνάρτηση μη γραμμικότητας, όπως είναι η Relu, παίρνουμε την έξοδο του *conv1*.

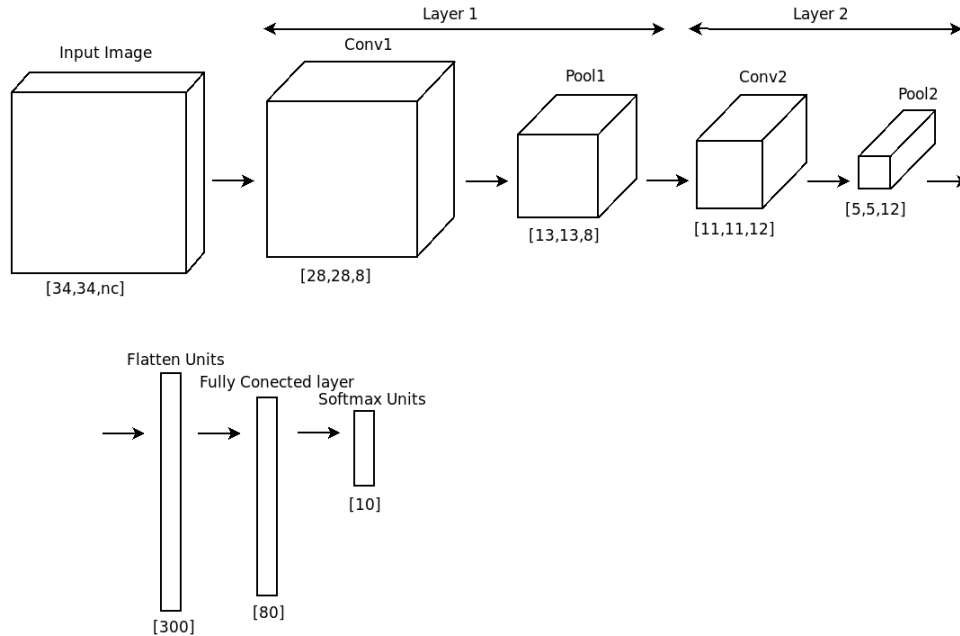
Στο επόμενο επίπεδο, εφαρμόζουμε μέγιστη συγκέντρωση, με διαστάσεις φίλτρου (4, 4) και βήμα δειγματοληψίας $s = 2$. Έτσι, οι διαστάσεις της εξόδου γίνονται (13, 13, 8), (ο αριθμός των καναλιών προκύπτει, όπως εξηγήθηκε) δημιουργώντας έτσι το επίπεδο *pool1*.

Με την σύμβαση η οποία ορίζει ένα νέο επίπεδο συνελικτικού δικτύου, μόνο στην περίπτωση που έχουμε νέα βάρη στο επίπεδο, ονομάζουμε τα *conv1* και *pool1* μαζί, ως *layer1* (το πρώτο επίπεδο) του δικτύου μας.

Στην συνέχεια εφαρμόζουμε ακόμη ένα συνελικτικό επίπεδο με τη χρήση 12 φίλτρων διαστάσεων (3, 3) και βήμα $s = 1$. Έτσι οι διαστάσεις της εξόδου του *conv2* προκύπτουν (11, 11, 12). Ακολουθεί πάλι ένα επίπεδο μέγιστης συγκέντρωσης (*pool2*) με $f = 3$ και $s = 2$. Οπότε η έξοδος του έχει διαστάσεις (5, 5, 12). Με την παραπάνω σύμβαση, ονομάζουμε το *conv2* και *pool2* μαζί ως δεύτερο επίπεδο (*layer2*) του δικτύου.

Έπειτα, αναδιαμορφώνουμε την τελευταία έξοδο, σε ένα διάνυσμα, διαστάσεων ίσο με $5 \times 5 \times 12 = 300$ το οποίο χρησιμοποιούμε ως είσοδο σε ένα πλήρες συνδεδεμένο επίπεδο με 80 μονάδες. Το επίπεδο αυτό έχει πίνακα βαρών w μεγέθους (80, 300) και πίνακα σταθερών b μεγέθους (80). Για πρόβλεψη πιθανοτήτων που να αθροίζουν στη μονάδα χρησιμοποιούμε την *softmax* συνάρτηση ενεργοποίησης κατά την έξοδο του νευρωνικού, η οποία έχει τύπο $softmax(x_i) = \exp(x_i) / \sum_j \exp(x_j)$.

Τελικά, έχουμε ένα διάνυσμα με ογδόντα αριθμούς, το οποίο τροφοδοτούμε σε μια μονάδα *softmax* με πλήθος εξόδων ίσο με το πλήθος των κλάσεων που θέλουμε να αναγνωρίσουμε.



Εικόνα 2.17: Απλή Αρχιτεκτονική Νευρωνικού Δικτύου.

Όπως φαίνεται στο Σχήμα 2.17, καθώς πηγαίνουμε βαθύτερα στο νευρωνικό δίκτυο, οι διαστάσεις ύψους (n_h) και πλάτους (n_w) μειώνονται, ενώ το πλήθος των καναλιών αυξάνεται.

2.2.7 Αρχιτεκτονικές Νευρωνικών Δικτύων Ευρείας Χρήσης

2.2.7.1 Το Νευρωνικό Δίκτυο VGG16

Αξιοσημείωτο γεγονός για το VGG16 είναι ότι αντί για πολλές υπερπαραμέτρους χρησιμοποιεί απλούστερο σχεδιασμό, στον οποίο τα συνελικτικά επίπεδα έχουν όλα διαστάσεις φίλτρων (3, 3) με βήμα ίσο με μονάδα και εφαρμόζουν όμοια συνέλιξη και τα επίπεδα μέγιστης συγκέντρωσης έχουν φίλτρα με διαστάσεις (2, 2) και βήμα δειγματοληψίας ίσο με 2.

Το δίκτυο VGG16 δέχεται ως είσοδο εικόνα διαστάσεων (224, 224, 3). Τα πρώτα δυο επίπεδα της αρχιτεκτονικής του είναι συνελικτικά επίπεδα με πλήθος φίλτρων ίσο με 64, οπότε παράγουν έξοδο διαστάσεων (224, 224, 64). Στην συνέχεια, ακολουθεί επίπεδο μέγιστης συγκέντρωσης, έτσι οι διαστάσεις των ενεργοποιήσεων μειώνονται στο μισό. Έπειτα, το VGG16 επαναλαμβάνει την σχεδιάσή του, έχοντας πάλι δύο στρώματα με 128 φίλτρα το καθένα και στη συνέχεια ένα στρώμα μέγιστης συγκέντρωσης. Έτσι, οι διαστάσεις των ενεργοποιήσεων γίνονται (56, 56, 128). Ακολουθούν 3 συνελικτικά επίπεδα με 256

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

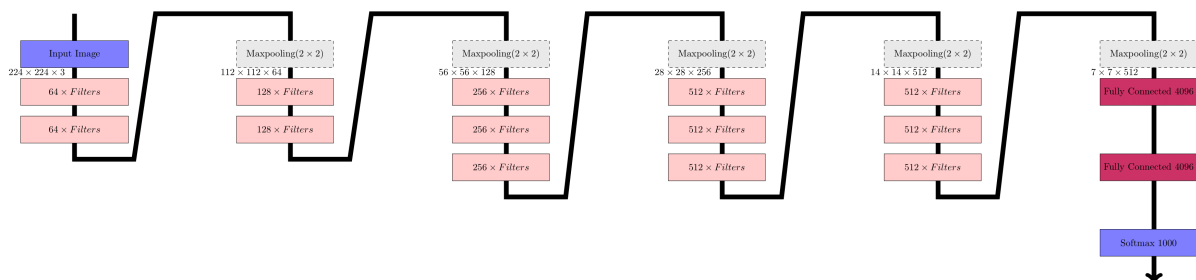
φίλτρα και ένα επίπεδο μέγιστης συγκέντρωσης και επαναλαμβάνεται το τελευταίο μέρος της σχεδίασης άλλη μια φορά με 512 φίλτρα. Οπότε, καταλήγει με ενεργοποιήσεις διαστάσεων (7, 7, 512).

Επειτα, η σχεδίαση συνδέει τις τελευταίες ενεργοποιήσεις με δύο πλήρως συνδεδεμένα επίπεδα, τα οποία αποτελούνται από 4096 κόμβους. Τέλος, οι ενεργοποιήσεις συνδέονται με μια συνάρτηση *softmax* 1000 εξόδων.

Και σε αυτήν την αρχιτεκτονική παρατηρούμε πως βαθύτερα στο νευρωνικό δίκτυο οι διαστάσεις των ενεργοποιήσεων μειώνονται (με συντελεστή δύο κάθε φορά) και ο αριθμός των καναλιών αυξάνεται (με συντελεστή δύο, όταν έχουμε νέα σειρά συνελικτικών επιπέδων).

Δύο από τα βασικά μειονεκτήματα του VGG16 είναι:

1. Πολλές παράμετροι βάρους που καθιστούν το μοντέλο αργό και στην εκμάθηση, αλλά και στην απόδοση των εκτιμήσεων.
2. Δεν μπορούμε να κάνουμε το μοντέλο πολύ πιο βαθύ, διότι βρισκόμαστε αντιμέτωποι με το πρόβλημα της εξαφάνισης των κλίσεων.



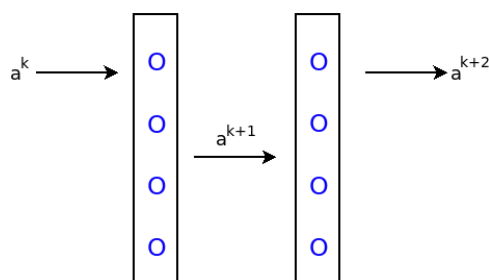
Εικόνα 2.18: Αρχιτεκτονική του VGG16.[1]

Ο αριθμός 16, στο όνομα του VGG, αναφέρεται στο γεγονός ότι αποτελείται από 16 επίπεδα (όπως φαίνεται στο Σχήμα 2.18) τα οποία έχουν νέα βάρη. Η αρχιτεκτονική του VGG16 το καθιστά αρκετά ελκυστικό από τους σχεδιαστές νευρωνικών δικτύων, διότι είναι αρκετά ομοιόμορφη και μπορεί να κατανοηθεί εύκολα. Για τον λόγο αυτό χρησιμοποιείται σε ένα μεγάλο πλήθος εφαρμογών νευρωνικών δικτύων.

2.2.7.2 Η αρχιτεκτονική ResNet

Νευρωνικά δίκτυα με πολλά επίπεδα είναι δύσκολο να εκπαιδευτούν λόγω των προβλημάτων τύπου εξαφάνισης ή έκρηξης των κλίσεων που προκύπτουν από τα γινόμενα των παραγώγων. Η αρχιτεκτονική ResNet χρησιμοποιεί παράλειψη των συνδέσεων ανάμεσα στα επίπεδα που επιτρέπει να πάρουμε τις ενεργοποιήσεις από ένα επίπεδο και να τις τροφοδοτήσουμε σε ένα άλλο επίπεδο βαθύτερα στο νευρωνικό δίκτυο. Με την τεχνική αυτή μπορούμε να εκπαιδεύσουμε νευρωνικά δίκτυα με περισσότερα επίπεδα. Δομικά στοιχεία των ResNets είναι τα υπολειπόμενα μπλοκ.

Ένα μη υπολειπόμενο μπλοκ δέχεται ως είσοδο a^k και παράγει έξοδο a^{k+1} , την οποία τροφοδοτεί στο επόμενο επίπεδο για τον υπολογισμό του a^{k+2} Σχήμα 2.19.

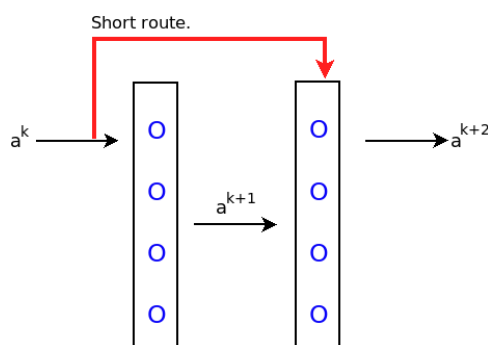


Εικόνα 2.19: Μη υπολειπόμενο μπλοκ.

Σε αυτό το μπλοκ έχουμε a^k ως είσοδο και στη συνέχεια εφαρμόζουμε μια γραμμική συνάρτηση σε αυτήν (για λογιστική παλινδρόμηση: $z^{k+1} = w^{k+1} \times a^k + b^{k+1}$ για τον υπολογισμό της εξόδου. Έπειτα, εφαρμόζουμε μια μη γραμμική συνάρτηση, υπολογίζοντας έτσι τις ενεργοποιήσεις a^{k+1} . Στο επόμενο βήμα, εφαρμόζουμε ξανά την ίδια γραμμική συνάρτηση, υπολογίζοντας με αυτόν τον τρόπο το z^{k+2} (για λογιστική παλινδρόμηση: $z^{k+2} = w^{k+2} \times a^{k+1} + b^{k+2}$) καθώς και πάλι μια μη γραμμική συνάρτηση για τον υπολογισμό των ενεργοποιήσεων a^{k+2} . Σε αυτό το σενάριο για να περάσουμε την πληροφορία, από το a^k στο a^{k+2} , πρέπει να παραμείνει ενεργή σε όλα τα βήματα του μπλοκ.

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

Σε ένα υπολειπόμενο μπλοκ προωθούμε προς τα εμπρός το a^k και το προσθέτουμε με το z^{k+2} , προτού εφαρμόσουμε την τελευταία μη γραμμική συνάρτηση Σχήμα 2.20.



Εικόνα 2.20: Υπολειπόμενο μπλοκ.

Οπότε τώρα, αντί η πληροφορία να ακολουθεί μόνο το κύριο μονοπάτι, έχει την δυνατότητα να ακολουθήσει την παράκαμψη για να πάει βαθύτερα στο νευρωνικό δίκτυο. Έτσι, με G μία μη γραμμική συνάρτηση, η τελευταία εξίσωση γίνεται $a^{k+2} = G(z^{k+2} + a^k)$. Ο τρόπος με τον οποίο φτιάχνουμε υπολειπόμενα νευρωνικά δίκτυα είναι η στοίβαξη πολλών υπολειπόμενων μπλοκ, το ένα ακόλουθο του άλλου.

Με την χρήση των υπολειπόμενων μπλοκ γίνεται εύκολο για τη συνάρτηση εκμάθησης να μάθει τις πληροφορίες από τα προηγούμενα στάδια, γιατί στη χειρότερη περίπτωση καταλήγει το a^{k+2} να είναι ίσο με a^k . Έτσι, προσθέτοντας νέα επίπεδα, δεν βλάπτεται η ικανότητα του νευρωνικού δικτύου ως προς την πρόβλεψη των κλάσεων.

Τέλος, αν και η αρχιτεκτονική ResNet βελτιώνει τα προβλήματα τύπου εξαφάνισης και έκρηξης των κλίσεων, καθιστά ακόμα πιο βαρύ το δίκτυο, έχοντας ως αποτέλεσμα να είναι πιο αργό και ως προς την εκπαίδευση και ως προς την απόκριση.

2.2.8 Αύξηση Δεδομένων Εκπαίδευσης

Στις περισσότερες εφαρμογές συνελκτικών νευρωνικών δικτύων καθίσταται η ανάγκη να χρησιμοποιηθούν περισσότερα δεδομένα εκπαίδευσης. Για αυτόν τον λόγο η αύξηση των δεδομένων είναι μια από τις τεχνικές που χρησιμοποιείται συχνά, με στόχο την βελτίωση της απόδοσης των συνελκτικών νευρωνικών δικτύων.

Η αύξηση των δεδομένων (για την εκπαίδευση των δικτύων όπου επιλύουν προβλήματα εφαρμογών μηχανικής όρασης), μπορεί να επιτευχθεί με αναστροφή, περιστροφή, εισαγωγή θορύβου και γενικά εφαρμόζοντας οποιαδήποτε διεργασία αλλαγής των εικόνων εκπαίδευσης.

2.2.9 Μεταφορά της Μάθησης Νευρωνικών Δικτύων

Αντί να εκπαιδεύσουμε ένα νευρωνικό δίκτυο από την αρχή, δηλαδή από κάποια τυχαία αρχικοποίηση των βαρών του, συχνά χρησιμοποιούμε βάρη που κάποιος άλλος έχει εκπαιδεύσει για την συγκεκριμένη αρχιτεκτονική του δικτύου σε διαφορετική εφαρμογή. Ο λόγος για αυτό πηγάζει από το γεγονός ότι τα δίκτυα με ήδη εκπαιδευμένα βάρη έχουν μάθει να αναγνωρίζουν σχέδια τα οποία χρειάζονται για την αναγνώριση των κλάσεων της εφαρμογής στην οποία εκπαιδεύτηκαν. Αυτά τα σχέδια όμως είναι πιθανό να ταιριάζουν και στην εφαρμογή που θέλουμε να υλοποιήσουμε. Για παράδειγμα, αν φτιάχνουμε ένα νευρωνικό δίκτυο το οποίο θέλουμε να αναγνωρίζει γάτες, τότε θα μπορούσαμε να κάνουμε αρχικοποίηση των βαρών του δικτύου με βάρη από ένα δίκτυο το οποίο αναγνωρίζει κάποιο άλλο είδος ζώου. Στην πράξη, έχει αποδειχθεί ότι ακόμα και αν δεν υπάρχει συσχέτιση ανάμεσα στις δύο εφαρμογές, η αρχικοποίηση με ήδη εκπαιδευμένα βάρη είναι μια καλή στρατηγική για την εκπαίδευση του δικτύου.

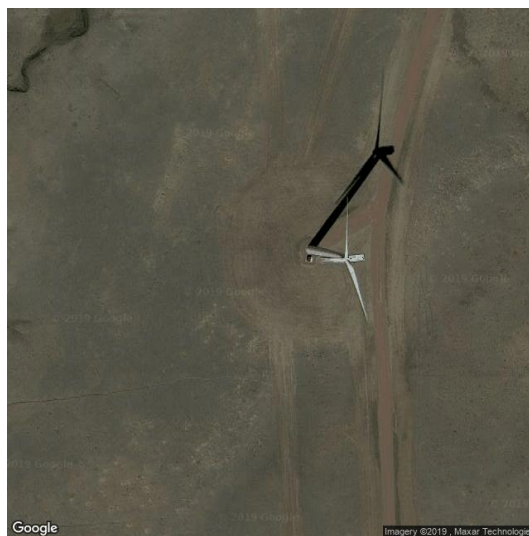
Η παραπάνω διαδικασία ονομάζεται Μεταφορά Μάθησης. Κατά την εκπαίδευση του μοντέλου ο σχεδιαστής επιλέγει εκείνα τα επίπεδα του νευρωνικού δικτύου τα οποία παραμένουν με τα αρχικοποιημένα βάρη (παγωμένα επίπεδα) και εκείνα που θα συμμετέχουν στην διαδικασία της εκπαίδευσης. Αυτή η επιλογή καθιστά την σχεδίαση πολύπλοκη διαδικασία, που η απόδοση της εξαρτάται τόσο από την κατανόηση των συνελικτικών νευρωνικών δικτύων, όσο και από το πλήθος διάφορων αρχιτεκτονικών.

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

2.2.10 Χάρτες Χαρακτηριστικών Συνελικτικού Νευρωνικού Δικτύου

Σε ένα συνελικτικό νευρωνικό δίκτυο οι μονάδες μέσα σε ένα εσωτερικό επίπεδο χωρίζονται σε χάρτες χαρακτηριστικών. Έτσι, οι μονάδες ενός χάρτη χαρακτηριστικών μοιράζονται τα βάρη των προηγούμενων επιπέδων για την εύρεση συγκεκριμένου χαρακτηριστικού.

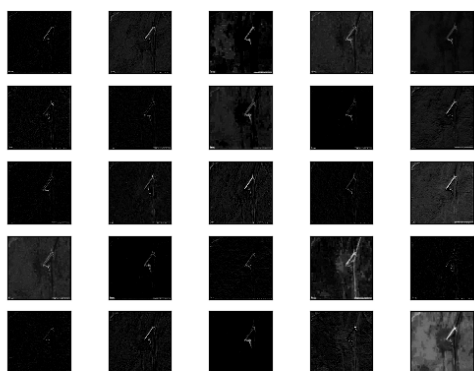
Χρησιμοποιώντας το νευρωνικό δίκτυο VGG16 εκπαιδευμένο σε προσαρμοσμένο σύνολο δεδομένων με εικόνες από ΑΠΕ (Ανανεώσιμων Πηγών Ενέργειας) και βάζοντας ως είσοδο την Εικόνα 2.21, απεικονίζουμε τους χάρτες χαρακτηριστικών από τρία συνελικτικά επίπεδα του νευρωνικού δικτύου στο Σχήμα 2.26.



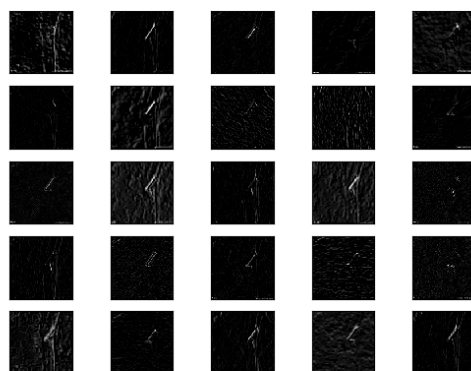
Όπως φαίνεται στις Εικόνες 2.22 και 2.23, οι χάρτες χαρακτηριστικών που βρίσκονται κοντά στην είσοδο του μοντέλου μοιάζουν αρκετά στην εικόνα εισόδου. Αυτό συμβαίνει επειδή οι χάρτες χαρακτηριστικών των πρώτων επιπέδων του δικτύου εξάγουν απλά χαρακτηριστικά όπως οριζόντιες και κάθετες γραμμές. Έτσι η απεικόνιση αυτών μοιάζει με εκείνη της εικόνας.

Για τα βαθύτερα επίπεδα του μοντέλου, οι χάρτες χαρακτηριστικών εντοπίζουν όλο και πιο σύνθετα στοιχεία της εικόνας. Έτσι αυτό που βλέπουμε ως άσπρο στις Εικόνες 2.24 και 2.25 είναι η εύρεση των αντίστοιχων χαρακτηριστικών από το δίκτυο.

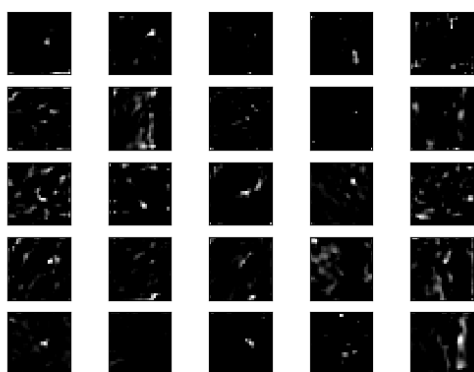
Εικόνα 2.21: Ανεμογεννήτρια από προσαρμοσμένο σύνολο δεδομένων.



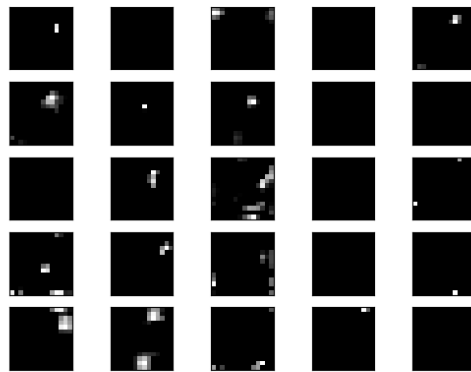
Εικόνα 2.22: Δεύτερο επίπεδο.



Εικόνα 2.23: Τέταρτο επίπεδο.



Εικόνα 2.24: Δέκατο επίπεδο.



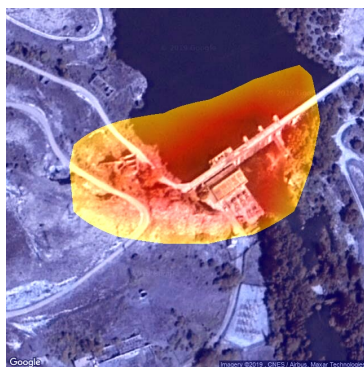
Εικόνα 2.25: Δέκατο τρίτο επίπεδο.

Εικόνα 2.26: Χάρτες χαρακτηριστικών από τα εσωτερικά επίπεδα του δικτύου.

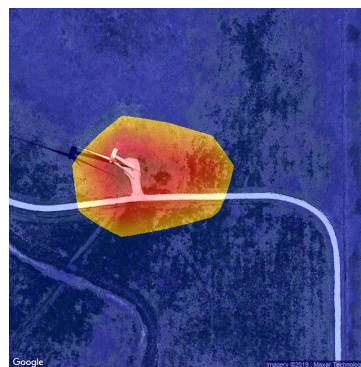
2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

2.2.11 Χάρτες Ενεργοποίησης των Κλάσεων

Η απεικόνιση των χαρτών ενεργοποίησης των κλάσεων είναι αυτό που βλέπουμε στις Εικόνες 2.27 και 2.28. Έχουμε επικαλύψει τις αρχικές δορυφορικές εικόνες με τους χάρτες θερμότητας (που παράγουμε με την τεχνική του χάρτη ενεργοποίησης των κλάσεων) και παρατηρούμε πως η περιοχή γύρω από το κύριο αντικείμενο που αναγνωρίζεται είναι θερμότερη από την υπόλοιπη εικόνα. Έτσι κατά μία έννοια, οι χάρτες ενεργοποίησης των κλάσεων βοηθούν στον εντοπισμό αντικειμένων, αφού εντοπίζεται η τοποθεσία τους στην εικόνα.



Εικόνα 2.27: Υδροηλεκτρική ενέργεια.



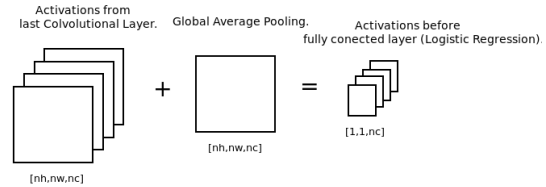
Εικόνα 2.28: Αιολική ενέργεια.

Εικόνα 2.29: Θερμικοί χάρτες από προσαρμοσμένο σύνολο δεδομένων.

Όπως έχει εξηγηθεί για νευρωνικά δίκτυα ταξινόμησης εικόνων, όπως το VGG16 ή το ResNet50, καθώς η εικόνα εισόδου μεταφέρεται βαθύτερα στο δίκτυο, παρόλο που οι διαστάσεις της συρρικνώνονται, ο αριθμός των χαρακτηριστικών που εντοπίζει το δίκτυο αυξάνεται.

Έτσι, για την εξαγωγή του χάρτη θερμότητας από εκπαιδευμένο δίκτυο ταξινόμησης εικόνων προσθέτουμε μετά το τελευταίο συνελικτικό επίπεδο, ένα επίπεδο ολικής μέσης συγκέντρωσης 2.30. Έτσι, δημιουργούμε μία τιμή για κάθε ένα χαρακτηριστικό του τελευταίου συνελικτικού επιπέδου. Στην συνέχεια τροφοδοτούμε ως είσοδο τις τιμές αυτές στο πλήρως συνδεδεμένο επίπεδο για να γίνει η πρόβλεψη της εξόδου.

Διαισθητικά, μπορούμε να σκεφτούμε κάθε χαρακτηριστικό του τελευταίου συνελικτικού επιπέδου ως ένα πιθανό κομμάτι στην εικόνα, για παράδειγμα, αν θέλουμε να αναγνωρίσουμε ένα ζώο, θα μπορούσαμε να έχουμε υψηλές τιμές στο χαρακτηριστικό που συσχετίζεται με τα μάτια ή



τα πόδια, αλλά τα χαρακτηριστικά τα οποία δεν Εικόνα 2.30: Ολική μέση συγκέντρωση. ταιριάζουν στην εικόνα, όπως ένας τροχός οχήματος, θα έχουν τιμές κοντά στο μηδέν. Από την άλλη, εάν η εικόνα εισόδου απεικονίζει ένα όχημα θα πάρουμε τα αντίθετα αποτελέσματα.

Ακόμη, για την κατανόηση του χάρτη ενεργοποίησης της κλάσης απαιτείται να σκεφτούμε προς τα πίσω την διαδρομή του δικτύου. Στην περίπτωση που το δίκτυο έχει βρει κάποιο χαρακτηριστικό σε μια εικόνα και η τιμή του χαρακτηριστικού είναι θετική μετά το επίπεδο της ολικής μέσης συγκέντρωσης, αυτό το χαρακτηριστικό υπάρχει κάπου στην εικόνα και πριν το επίπεδο της ολικής μέσης συγκέντρωσης είχαμε πληροφορία για την τοποθεσία του χαρακτηριστικού στην εικόνα.

Μια ακόμη έννοια για την έμπνευση του χάρτη ενεργοποίησης της κλάσης προέρχεται από τον τρόπο με τον οποίο λειτουργεί η λογιστική παλινδρόμηση. Για τιμές ενεργοποιήσεων μεγαλύτερες από μηδέν και για θετικά βάρη τα οποία συνδέονται με την εκάστοτε κλάση έχουμε χαρακτηριστικά τα οποία υπάρχουν στην αντίστοιχη εικόνα εισόδου. Αν τα βάρη είναι κοντά στο μηδέν, τα χαρακτηριστικά δεν έχουν καμία συσχέτιση με την κλάση, ενώ αν είναι αρνητικά, αυτό σημαίνει ότι είναι λιγότερο πιθανό να υπάρχουν αυτά τα χαρακτηριστικά στην εικόνα.

Κατά την δημιουργία του χάρτη θερμότητας της εικόνας, θεωρούμε σημαντικά μόνο εκείνα τα βάρη που προέρχονται από την μεγαλύτερη (ή τις μεγαλύτερες) πρόβλεψη των κλάσεων. Έτσι, όταν προβλέπουμε την ύπαρξη ενός ζώου στην εικόνα χρησιμοποιούνται εκείνα τα βάρη που αντιστοιχούν στην κλάση "ζώο".

Τα χαρακτηριστικά πριν την εφαρμογή του επιπέδου ολικής μέσης συγκέντρωσης έχουν διαστάσεις (7, 7), έτσι πολλαπλασιάζοντάς τα με τα βάρη από την λογιστική παλινδρόμηση υπολογίζεται η σημασία του καθενός. Αυτός ο υπολογισμός είναι το εξωτερικό γινόμενο των βαρών και των τελευταίων ενεργοποιήσεων. Η έξοδος αυτού του υπολογισμού είναι μία εικόνα διαστάσεων (7, 7) και με την αλλαγή μεγέθους της εικόνας αυτής, ώστε να είναι ίση με το αρχικό μέγεθος της εικόνας εισόδου, έχουμε το θερμικό χάρτη της προβλεπόμενης κλάσης.

2.3 Η Μετρική F-score

Η F-score μετρική χρησιμοποιείται συνήθως στην ανάλυση των προβλέψεων. Έτσι, στην περίπτωση που έχουμε ένα δυαδικό πρόβλημα ταξινόμησης, τότε προβλέπουμε αποτελέσματα είτε θετικά είτε αρνητικά. Αυτό μας οδηγεί σε τέσσερα πιθανά σενάρια προβλέψεων.

- Ορθά αρνητική πρόβλεψη (TN): Σε αυτό το σενάριο έχουμε αρνητική περίπτωση παραδείγματος, καθώς και αρνητική πρόβλεψη.
- Ορθά θετική πρόβλεψη (TP): Σε αυτό το σενάριο έχουμε θετική περίπτωση παραδείγματος, καθώς και θετική πρόβλεψη.
- Λανθασμένα αρνητική πρόβλεψη (FN): Σε αυτό το σενάριο έχουμε θετική περίπτωση παραδείγματος, αλλά αρνητική πρόβλεψη.
- Λανθασμένα θετική πρόβλεψη (FP): Σε αυτό το σενάριο έχουμε αρνητική περίπτωση παραδείγματος, αλλά θετική πρόβλεψη.

Ο πίνακας σύγχυσης προσδίδει τον αριθμό των τεσσάρων παραπάνω σεναρίων σε ένα πρόβλημα ταξινόμησης.

	Προβλεπόμενο αρνητικό	Προβλεπόμενο θετικό
αρνητικό	TN	FP
θετικό	FN	TP

Για την εξήγηση του F-score, πρώτα πρέπει να αναλυθούν οι έννοιες των Precision και Recall.

- Precision είναι το σύνολο των ορθά θετικών προβλέψεων διαιρεμένο με το σύνολο όλων των θετικών προβλέψεων και υπολογίζεται από την εξίσωση: $Precision = TP / (TP + FP)$.
- Recall είναι το σύνολο των ορθά θετικών προβλέψεων διαιρεμένο με το σύνολο όλων των πραγματικά θετικών παραδειγμάτων και υπολογίζεται από την εξίσωση: $Recall = TP / (TP + FN)$.

Έτσι, η μετρική F-score αποτελεί το αρμονικό μέσο των Precision και Recall και υπολογίζεται από τον τύπο:

$$F = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Οπότε, το σύνολο τιμών του F-score είναι $[0, 1]$ και όσο μεγαλύτερη τιμή έχει τόσο καλύτερη είναι η ταξινόμηση της κλάσης.

2. ΑΝΑΣΚΟΠΗΣΗ ΘΕΩΡΙΑΣ

Κεφάλαιο 3

Διατύπωση του Προβλήματος

3.1 Ανάγκη Κατανόησης των Νευρωνικών Δικτύων

Οι αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων ποικίλουν, και επίσης έχει αποδειχθεί αναμφισβήτητα πως η χρήση τους καθίσταται απαραίτητη για την επίλυση πολλών προβλημάτων. Αν και η χρήση βαθιών νευρωνικών δικτύων είναι πλέον εργαλείο σε πολλά χέρια, για πολλούς παραμένει να είναι ένα μαύρο κουτί, το οποίο χρησιμοποιείται λόγω των καταπληκτικών αποτελεσμάτων που επιφέρει. Έτσι, για την καλύτερη χρήση τους, καθώς και για την εξέλιξή τους, είναι απαραίτητη η διερεύνηση και κατανόηση της λειτουργίας τους.

Σίγουρα σε αυτήν την χρονική περίοδο που διανύουμε τα νευρωνικά δίκτυα είναι στο προσκήνιο της έρευνας, με αποτέλεσμα να προκύπτουν συνεχώς νέες αρχιτεκτονικές με όλο και μεγαλύτερες αποχρίσεις. Όμως, εκτός από την διερεύνηση νέων αρχιτεκτονικών, υπάρχει η ανάγκη κατανόησης του τρόπου λειτουργίας των παλαιότερων. Αυτό επιτυγχάνεται, εφευρίσκοντας τεχνικές που δίνουν προσοχή τόσο στα σύνολα δεδομένων που χρησιμοποιούνται για την εκπαίδευση των νευρωνικών δικτύων, όσο και στα εσωτερικά στοιχεία που αποτελούν το ίδιο το δίκτυο (εσωτερικά επίπεδα των δικτύων, οι χάρτες χαρακτηριστικών, αλλά και οι θερμικοί χάρτες που παράγονται από διάφορες μεθόδους, όπως οι χάρτες ενεργοποίησης των κλάσεων).

Έτσι, για να χτιστεί εμπιστοσύνη στο κατά πόσο είναι ορθή η επιλογή νευρωνικού δικτύου για την επίλυση ενός προβλήματος, καθώς και για να οριστεί η κατεύθυνση που πρέπει να ακολουθήσουν οι ερευνητές για την εξέλιξη ενός νευρωνικού δικτύου, πρέπει να διερευνηθούν οι λόγοι πίσω από τον τρόπο με τον οποίο γίνονται οι προβλέψεις στα εν λόγω δίκτυα.

3. ΔΙΑΤΥΠΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

3.2 Μέθοδοι Διεύρυνσης του Συνόλου Δεδομένων

Αναμφίβολα, ο σχεδιασμός της αρχιτεκτονικής ενός νευρωνικού δικτύου κατέχει κύριο ρόλο για την συμπεριφορά του. Όμως, το σημαντικότερο κομμάτι της ποιότητας των αποτελεσμάτων ενός νευρωνικού δικτύου, πηγάζει πάντα από το σύνολο δεδομένων στο οποίο εκπαιδεύτηκε.

Καλή αρχιτεκτονική σχεδίαση είναι εκείνη που έχει καλύτερη απόκριση και ως προς τον χρόνο, αλλά και ως προς την ακρίβεια, σε σχέση με μια άλλη αρχιτεκτονική, δεδομένου ότι και οι δύο έχουν εκπαιδευτεί στο ίδιο σύνολο δεδομένων. Όμως, αν υπάρχει η δυνατότητα να αυξήσουμε τα δεδομένα εκπαίδευσης στην χειρότερη αρχιτεκτονική, τότε εύκολα μπορεί να ξεπεράσει την απόδοση της καλύτερης.

Από τα παραπάνω γίνεται εμφανές πόσο σημαντικό είναι να δοθεί έμφαση στην ποιοτική αύξηση του συνόλου δεδομένων εκπαίδευσης ενός νευρωνικού δικτύου, έτσι ώστε να επιτευχθεί το μέγιστο δυνατό αποτέλεσμα. Αυτό επιτυγχάνεται με την δημιουργία αυτοματοποιημένων προγραμματιστικών εργαλείων, που έχουν ως στόχο να αντλήσουν δεδομένα, τα οποία σχετίζονται με το ζήτημα που μας ενδιαφέρει, εξαντλώντας τις υπάρχουσες δυνατότητες.

3.3 Σύνδεση Νευρωνικών Δικτύων και ΑΠΕ

Σε ότι αφορά την δημιουργία συνόλων δεδομένων δορυφορικών εικόνων τα οποία προορίζεται για εκπαίδευση νευρωνικών δικτύων, φαίνεται να παρουσιάζουν σημαντικό ενδιαφέρον εκείνες οι κατηγορίες που συμπεριλαμβάνουν δορυφορικές εικόνες που απεικονίζουν γεννήτριες ΑΠΕ.

Κατά το παρελθόν, έχουν δημιουργηθεί νευρωνικά δίκτυα όχι μόνο για τον εντοπισμό γεννητριών ΑΠΕ από δορυφορικές εικόνες, αλλά και για την εύρεση του ιδανικού φυσικού περιβάλλοντος στο ζήτημα δημιουργίας νέων εγκαταστάσεων. Αυτά τα δίκτυα αφενός προσφέρουν θετικά στην οικονομική επένδυση των έργων και ως προς το συμφέρον των επενδυτών, αφετέρου δίνουν βέλτιστες λύσεις οι οποίες είναι προς όφελος του περιβάλλοντος.

Με βάση τα παραπάνω, ορίζεται η ανάγκη της περαιτέρω διερεύνησης των δυνατοτήτων που μπορούν να προσφέρουν τα σύνολα δεδομένων από γεννήτριες ΑΠΕ, καθώς και απαιτείται η εφεύρεση νέων μεθόδων, με τις οποίες μπορούμε να εκμεταλλευθούμε με καλύτερο τρόπο τέτοια σύνολα δεδομένων.

3.4 Σχετικές Εργασίες

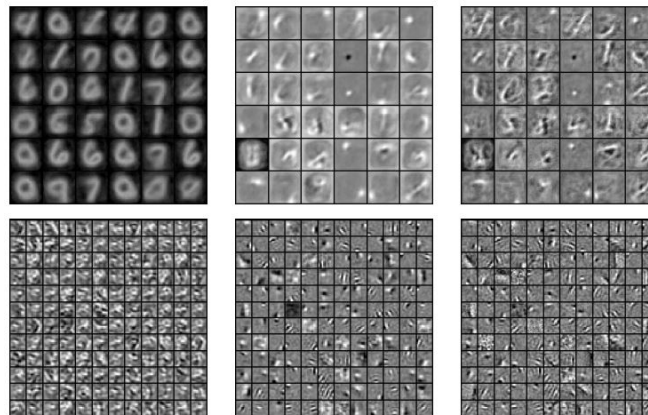
3.4.1 Μέθοδοι Απεικόνισης των Χαρακτηριστικών

Visualizing Higher-Layer Features of a Deep Network

Το 2009 οι Dumitru Erhan, Yoshua Bengio, Aaron Courville και Pascal Vincent δημοσίευσαν την Εργασία [2] η οποία είχε ως στόχο να βρει καλές και ποιοτικές ερμηνείες των χαρακτηριστικών που αφορούν συνελικτικά νευρωνικά δίκτυα.

Έτσι, με τις τεχνικές που όρισαν, έδωσαν στους ερευνητές την δυνατότητα να κατανοήσουν τον τρόπο με τον οποίο λειτουργούν τα βαθιά νευρωνικά δίκτυα.

Για να επιτευχθεί ο στόχος τους, παρουσίασαν δύο νέες τεχνικές, την μεγιστοποίηση των ενεργοποιήσεων και την μεμονωμένη χρήση παραδειγμάτων από ένα μόνο επίπεδο δικτύου, καθώς και μια παλιότερη μέθοδο, τον γραμμικό συνδυασμό των φίλτρων προηγούμενων επιπέδων.



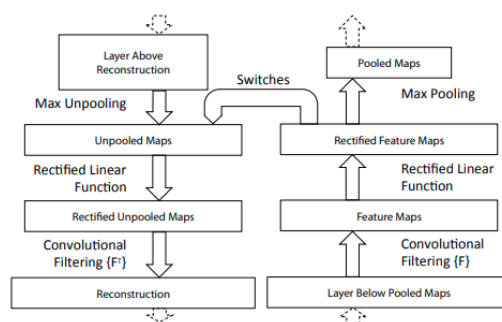
Εικόνα 3.1: Χάρτες Χαρακτηριστικών του δευτέρου επιπέδου των δύο δικτύων που χρησιμοποιήθηκαν. Η εικόνα προέρχεται από την [2].

Στην Εικόνα 3.1 απεικονίζονται οι χάρτες χαρακτηριστικών, του δευτέρου επιπέδου, των δύο δικτύων που χρησιμοποίησαν κατά την εργασία τους, με χρήση των τριών τεχνικών που προτείνουν. Αριστερά, απεικονίζονται με την τεχνική της μεμονωμένης χρήσης παραδειγμάτων από μόνο ένα επίπεδο δικτύου, δεξιά με την τεχνική της μεγιστοποίησης των

3. ΔΙΑΤΥΠΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

ενεργοποιήσεων του επιπέδου που μας ενδιαφέρει και στην μέση με την μέθοδο του γραμμικού συνδυασμού των φίλτρων προηγούμενων επιπέδων.

Visualizing and Understanding Convolutional Networks



Εικόνα 3.2: Επίπεδο του deconvnet το οποίο είναι ενωμένο παράλληλα με τον αντίστοιχο επίπεδο του νευρωνικού δικτύου. Με αυτό το επίπεδο ανακατασκευάζονται τα χαρακτηριστικά του προηγούμενου επιπέδου. Η εικόνα προέρχεται από την [3].

Μια αντίστοιχη δουλειά η οποία συνεισφέρει στην κατανόηση της λειτουργίας των συνελικτικών νευρωνικών δικτύων είναι η Εργασία [3]. Σε αυτήν την δουλειά χρησιμοποιήθηκαν πλήρως συνελικτικά νευρωνικά μοντέλα με στόχο να ταξινομηθούν εικόνες εισόδου σε κάποιο ορισμένο πλήθος κλάσεων.

Τα επίπεδα των μοντέλων τους αποτελούνται από τέσσερα βασικά στοιχεία. Πρώτον, από την συνέλιξη που προκύπτει από τον προηγούμενο επίπεδο σε συνδυασμό με τα εκπαιδευμένα φίλτρα. Δεύτερον, από την συνάρτηση ενεργοποίησης ReLu για τον υπολογισμό των ενεργοποιήσεων. Τρίτον, από ένα επίπεδο μέγιστης συγκέντρωσης. Και τέλος, από μια τοπική λειτουργία, η οποία εξομαλύνει τις αποκρίσεις των χαρτών των χαρακτηριστικών. Ως συνάρτηση απώλειας, χρησιμοποιούν την cross-entropy, καθώς και για την ενημέρωση των βαρών των δικτύων τους χρησιμοποίησαν τον αλγόριθμο stochastic gradient descent.

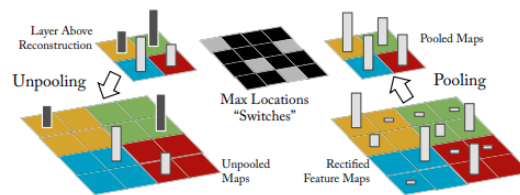
Στην Εικόνα 3.2 παρουσιάζεται η αντιστοιχία ενός επιπέδου του deconvnet με ένα επίπεδο του μοντέλου.

Για να επιτευχθεί η κατανόηση της λειτουργίας των μοντέλων, δίνουν έμφαση στην ερμηνεία των χαρτών των χαρακτηριστικών που προέκυψαν από τα ενδιάμεσα επίπεδα των μοντέλων.

Η παραπάνω ερμηνεία γίνεται με τον εντοπισμό των μοτίβων εισόδου που προκάλεσαν μια δεδομένη ενεργοποίηση στους χάρτες χαρακτηριστικών, χρησιμοποιώντας ένα deconvolutional δίκτυο ή όπως το ονομάζουν ένα deconvnet.

Το deconvnet, αντί να αντιστοιχεί τα εικονοστοιχεία σε χαρακτηριστικά, όπως ένα συνελκτικό νευρωνικό δίκτυο, κάνει το αντίθετο.

Για να γίνει ο έλεγχος μιας ενεργοποίησης των μοντέλων, αρχικά μηδενίζονται όλες οι άλλες ενεργοποιήσεις και έπειτα περνούν τους χάρτες χαρακτηριστικών στο αντίστοιχο επίπεδο του deconvnet. Έτσι, με την διαδικασία της αντίθετης συγκέντρωσης (Εικόνα 3.3), της εφαρμογής της συνάρτησης ReLu και των φίλτρων επιτυγχάνεται η ανακατασκευή της εισόδου από το προηγούμενο επίπεδο, το οποίο προκάλεσε την ενεργοποίηση που μας ενδιαφέρει. Επαναλαμβάνοντας την διαδικασία αυτή μέχρι να βρεθούμε στο επίπεδο της εισόδου, θα έχουμε ως αποτέλεσμα μια εικόνα ενεργοποιήσεων με διαστάσεις ίδιες με εκείνες της εικόνας εισόδου.



Εικόνα 3.3: Μια αναπαράσταση της αντίθετης συγκέντρωσης. Ο προσδιορισμός της τοποθεσίας των μεγίστων γίνεται με τους διακόπτες. Η εικόνα προέρχεται από την [3].

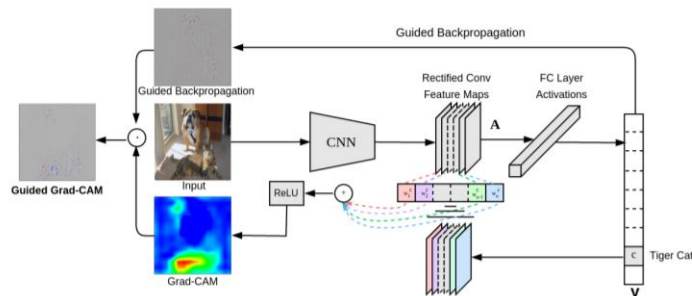
3. ΔΙΑΤΥΠΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

3.4.2 Μέθοδοι Επιβεβαίωσης των Προβλέψεων

Grad-CAM: "Why did you say that?"

Με στόχο την καλύτερη κατανόηση του τρόπου με τον οποίο ταξινομούν τα συνελικτικά νευρωνικά δίκτυα, το 2017 οι Ramprasaath R.Selvaraju, Abhishek Das, Devi Parikh, Ramakrishna Vedantam, Dhruv Batra και Michael Cogswell παρουσιάζουν μια τεχνική με την οποία γίνεται απεικόνιση των περιοχών που ένα νευρωνικό δίνει έμφαση κατά την ταξινόμηση. Η προσέγγισή τους ονομάζεται Guided Grad-CAM. Με την Εργασία [4] πέτυχαν την δημιουργία θερμικών χαρτών, οι οποίοι έχουν και υψηλή ανάλυση, αλλά και μεγάλη ακρίβεια.

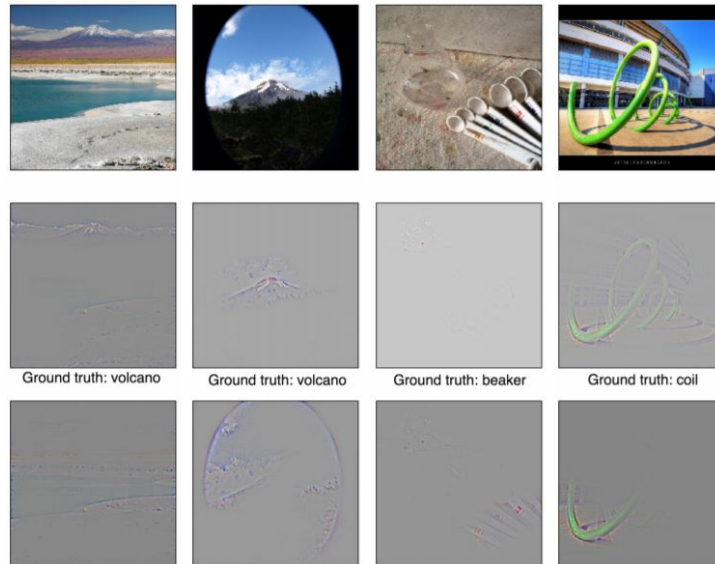
Στην εργασία τους διατυπώνουν ότι ως καλή επεξήγηση μιας απεικόνισης ορίζεται η καλή διακριτότητα της εικόνας, καθώς και η καλή ανάλυση της. Με αυτές τις προϋποθέσεις κατάφεραν να ανταπεξέλθουν με την μέθοδό τους.



Εικόνα 3.4: Γραφική Αναπαράσταση της μεθόδου Guided Grad-CAM. Η εικόνα προέρχεται από την [4].

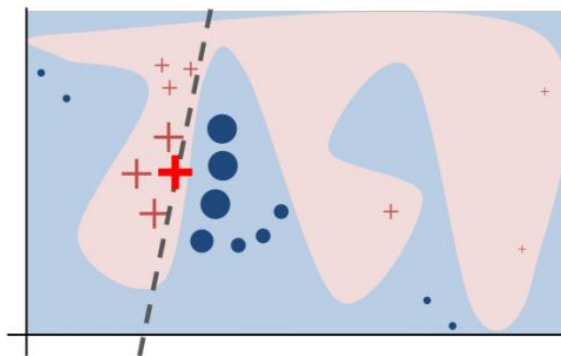
Όπως φαίνεται στην Εικόνα 3.4 δίνεται μια εικόνα, καθώς και η κατηγορία της, και στην συνέχεια εξάγονται οι χάρτες χαρακτηριστικών πριν την συνάρτηση softmax. Έπειτα, όλες οι κλίσεις τίθενται ως μηδενικές, πλην εκείνης της κλάσης που μας ενδιαφέρει η οποία τίθεται ως μονάδα. Στην συνέχεια, το παραγόμενο σήμα ανατροφοδοτείται και με την χρήση της συνάρτησης ReLU δημιουργείται ο θερμικός χάρτης της κλάσης που μας ενδιαφέρει. Τέλος, πολλαπλασιάζοντας τον θερμικό χάρτη με την αρχική εικόνα εισόδου, παράγεται μια απεικόνιση υψηλής ανάλυσης και διάκρισης.

Στην Εικόνα 3.5 βλέπουμε παραδείγματα της μεθόδου Guided Grad-CAM. Όπως φαίνεται, με την μέθοδο εντοπίζονται εκείνα τα σημεία όπου τα νευρωνικά δίνουν έμφαση κατά την ταξινόμηση.



Εικόνα 3.5: Παραδείγματα της μεθόδου Guided Grad CAM. Η εικόνα προέρχεται από την [4].

“Why Should I Trust You?”.



Εικόνα 3.6: Μεθόδος LIME. Η εικόνα προέρχεται από την [5].

Το 2016 οι Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin δημοσίευσαν την

3. ΔΙΑΤΥΠΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

εργασία με τίτλο “Why Should I Trust You?” [5] με την οποία προτείνουν την μέθοδο LIME. Η μέθοδος αυτή εξηγεί τον τρόπο με τον οποίο ταξινομεί οποιοσδήποτε ταξινομητής. Ακόμη, προτείνουν μια τεχνική κατανόησης των δικτύων, δίνοντας έμφαση σε μεμονωμένα υπο-προβλήματα της ταξινόμησης. Τέλος, καθορίζουν αν ο χρήστης πρέπει να εμπιστεύεται έναν ταξινομητή για την επίλυση του προβλήματος της ταξινόμησης που τον ενδιαφέρει.

Στην Εικόνα 3.6 βλέπουμε μια απεικόνιση του τρόπου λειτουργίας της μεθόδου LIME. Η μέθοδος αυτή έχει άγνοια για την συνάρτηση απώλειας που χρησιμοποιεί το μοντέλο κατά την διαδικασία της πρόβλεψης. Όμως με γραμμική εφαρμογή, δίνει μεγαλύτερη έμφαση σε εκείνες τις περιπτώσεις που βρίσκονται σε μικρότερη απόσταση από την διακεκομμένη γραμμή και μικρότερη σε εκείνες που βρίσκονται σε μεγαλύτερη απόσταση. Έτσι, γίνεται η εκτίμηση για τον αν οι προβλέψεις του μοντέλου εμπνέουν εμπιστοσύνη.



Εικόνα 3.7: Απεικόνιση των τριών μεγαλύτερων προβλέψεων στη μέθοδο LIME. Η εικόνα προέρχεται από την [5].

Μια ακόμη μέθοδος που προτείνεται από την εργασία αυτή είναι η διερεύνηση και απεικόνιση των μεγαλύτερων ενεργοποιήσεων της εξόδου κατά την ταξινόμηση ενός συνελικτικού νευρωνικού δικτύου. Στην Εικόνα 3.7 παρουσιάζονται οι τρεις μεγαλύτερες προβλέψεις που προέκυψαν κατά την ταξινόμηση εικόνας που έγινε με το νευρωνικό δίκτυο Inception της Google.

Κεφάλαιο 4

Η Προσέγγισή μας

4.1 Ανάλυση Συστήματος Συντεταγμένων

Η Google και το OpenStreetMap χρησιμοποιούν το ακόλουθο σύστημα συντεταγμένων:

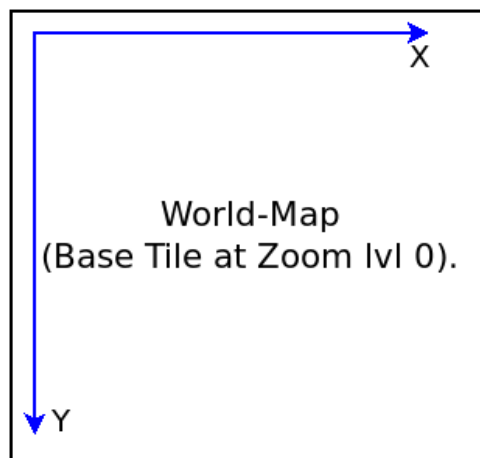
- Χρησιμοποιώντας το γεωδαιτικό σύστημα WGS84, ορίζουν συντεταγμένες γεωγραφικού πλάτους και γεωγραφικού μήκους, για κάθε σημείο του κόσμου ξεχωριστά.
- Ακόμη, ορίζουν τις παγκόσμιες συντεταγμένες οι οποίες αναφέρονται σε κάθε σημείο του χάρτη ξεχωριστά.
- Για τα διαφορετικά επίπεδα μεγέθυνσης του χάρτη ορίζονται οι συντεταγμένες εικονοστοιχείων, οι οποίες αναφέρονται σε ένα μοναδικό εικονοστοιχείο του χάρτη, για συγκεκριμένο επίπεδο μεγέθυνσης.
- Τέλος, ορίζονται οι συντεταγμένες πλακιδίων, οι οποίες αναφέρονται σε ένα μοναδικό πλακάκι του χάρτη για συγκεκριμένο επίπεδο μεγέθυνσης.

Παγκόσμιες συντεταγμένες:

Αρχικά, για την αντιστοίχιση των σημείων του κόσμου σε έναν χάρτη πρέπει πρώτα να μετατραπούν οι συντεταγμένες γεωγραφικού πλάτους και γεωγραφικού μήκους σε παγκόσμιες συντεταγμένες.

Επί της ουσίας, οι παγκόσμιες συντεταγμένες είναι ίδιες με τις συντεταγμένες εικονοστοιχείων, όταν το επίπεδο μεγέθυνσης του χάρτη είναι ίσο με το μηδέν (Εικόνα [4.1](#)).

4. Η ΠΡΟΣΕΓΓΙΣΗ ΜΑΣ



Εικόνα 4.1: Παγκόσμιες συντεταγμένες.

Οπότε, από τον ορισμό τους βλέπουμε ότι οι παγκόσμιες συντεταγμένες είναι ανεξάρτητες από το επίπεδο μεγέθυνσης.

Αν οι διαστάσεις ενός πλακιδίου είναι (x_0, y_0) , τότε οι παγκόσμιες συντεταγμένες έχουν σύνολο τιμών ίσο με $x \in [0, x_0)$, $y \in [0, y_0)$. Αυτό προκύπτει, διότι στο επίπεδο μηδέν έχουμε ένα πλακάκι το οποίο αντιστοιχεί σε όλο τον χάρτη. Για την δική μας υλοποίηση, το μέγεθος του πλακιδίου είναι σταθερό και έχει διαστάσεις $(256, 256)$.

Συντεταγμένες εικονοστοιχείων:

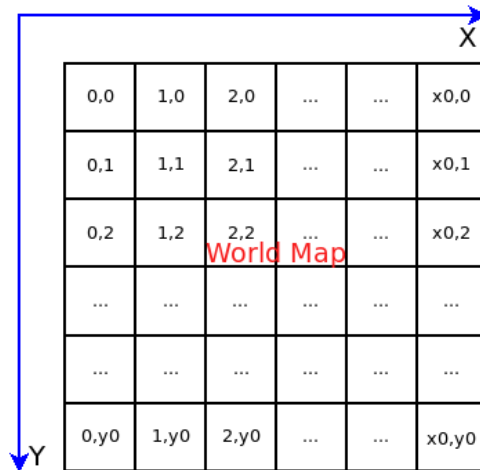
Σε δεδομένο επίπεδο μεγέθυνσης, οι συντεταγμένες εικονοστοιχείων αναφέρονται σε ένα μοναδικό εικονοστοιχείο του χάρτη και υπολογίζονται από τον τύπο: $PixelCoord = WorldCoord * 2^{ZoomLevel}$.

Έτσι, αφού οι παγκόσμιες συντεταγμένες έχουν οριστεί με βάση το βασικό πλακάκι (επίπεδο μεγέθυνσης ίσο με το μηδέν), το οποίο είναι ίσο με το μέγεθος του πλακιδίου, μπορούμε να βρούμε σε πιο εικονοστοιχείο του χάρτη (σε ένα συγκεκριμένο επίπεδο μεγέθυνσης) αντιστοιχεί ένα ζεύγος συντεταγμένων εικονοστοιχείων, από το ακέραιο μέρος τους.

Συντεταγμένες πλακιδίων:

Οι χάρτες της Google και του OpenStreetMap, για ένα συγκεκριμένο επίπεδο μεγέθυν-

σης, είναι διαχωρισμένοι σε $2^{ZoomLevel} * 2^{ZoomLevel}$ πλακάκια. Όπως φαίνεται στην Εικόνα 4.2, τα πλακάκια αριθμούνται από τα αριστερά προς τα δεξιά για τον άξονα X και από πάνω προς τα κάτω για τον άξονα Y .



Εικόνα 4.2: Αναπαράσταση συντεταγμένων των πλακιδίων του χάρτη.

Για τον υπολογισμό των συντεταγμένων πλακιδίου σε ένα συγκεκριμένο επίπεδο μεγέθυνσης διαιρούμε τις συντεταγμένες εικονοστοιχείου με το μέγεθος του πλακιδίου και στην συνέχεια παίρνουμε το ακέραιο μέρος της διαίρεσης.

4.2 Η Εφαρμογή Maps Static

Η εφαρμογή Maps Static της Google επιστρέφει μια δορυφορική εικόνα ως απάντηση σε ένα HTTP αίτημα. Για κάθε εικόνα μπορούμε να ορίσουμε την τοποθεσία που βρίσκεται στον χάρτη, το μέγεθος της εικόνας (σε εικονοστοιχεία), τον τύπο της εικόνας, καθώς και το επίπεδο μεγέθυνσης του χάρτη.

Για να χρησιμοποιήσουμε την εφαρμογή Maps Static, έπρεπε να ενεργοποιήσουμε ένα προσωπικό κλειδί στο Google Cloud, το οποίο στέλνουμε μαζί με κάθε αίτημα, ώστε να μας επιτραπεί η πρόσβαση στα δεδομένα.

Με την ενεργοποίηση του Google Cloud λογαριασμού, μας παρέχεται ένα αρχικό χρηματικό πόσο, το οποίο ήταν αρκετό για την διεκπεραίωση αυτής της εργασίας. Η χρέωση για περαιτέρω χρήση της εφαρμογής Maps Static φαίνεται στην Εικόνα 4.3.

4. Η ΠΡΟΣΕΓΓΙΣΗ ΜΑΣ

0–100,000	100,001–500,000
0.002 USD per each (2.00 USD per 1000)	0.0016 USD per each (1.60 USD per 1000)

Εικόνα 4.3: Χρέωση για την εφαρμογή Static Map.

Για την υλοποίηση αυτής της διπλωματικής εργασίας χρησιμοποιήσαμε την Maps Static εφαρμογή κατά την δημιουργία ενός προγραμματιστικού εργαλείου, με το οποίο εξάγουμε με δυναμικό τρόπο προσαρμοσμένα σύνολα δεδομένων δορυφορικών εικόνων για την χρήση σε εφαρμογές Μηχανικής Όρασης.

4.3 Overpass API, QL και OSM Tile Calculator

Overpass API:

Η εφαρμογή Overpass δέχεται μια εντολή για ανάγνωση στην οποία απαντά με επιλεγμένα κομμάτια δεδομένων του OpenStreetMap. Οπότε, στέλνοντας ο χρήστης ένα ερώτημα στην εφαρμογή, λαμβάνει δεδομένα του OpenStreetMap.

Κύριο αρνητικό στοιχείο της εφαρμογής Overpass είναι πως ανταποκρίνεται σε λίγα ερωτήματα που θα τεθούν από έναν χρήστη, σε μια συγκεκριμένη χρονική περίοδο. Τα ερωτήματα πρέπει να είναι στην μορφή της QL (Query Language).

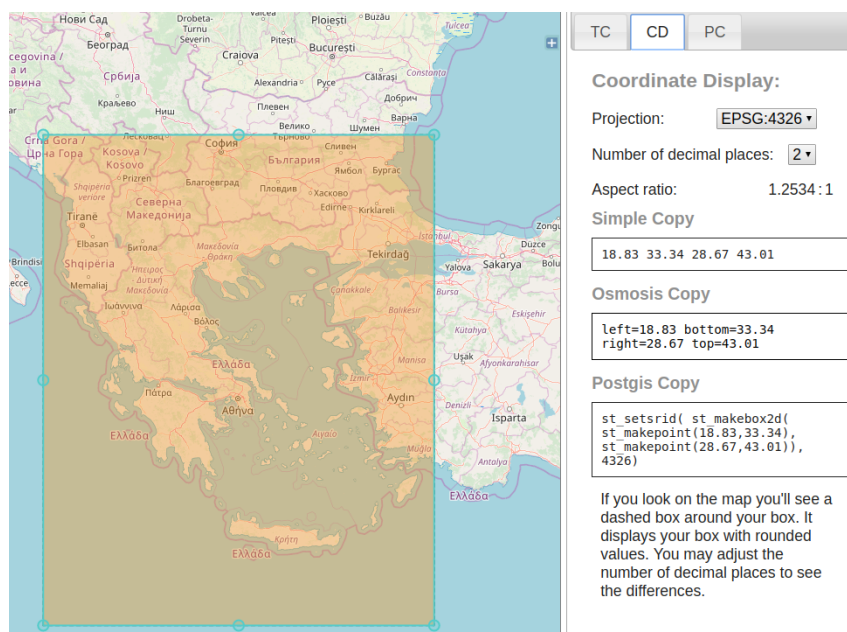
Overpass Query Language:

Με την QL ζητάμε δεδομένα από τη βάση δεδομένων του OSM, που βρίσκονται μέσα σε ένα οριοθετημένο χώρο (bbox), το οποίο αντιστοιχεί σε ένα μέρος του παγκόσμιου χάρτη. Το bbox ορίζεται με τέσσερις συντεταγμένες [νότια, δυτικά, βόρεια, ανατολικά]. Η εφαρμογή Overpass υποστηρίζει αρχεία εξόδου σε μορφή CSV (αρχεία διαχωρισμένα με κόμμα).

OSM Tile Calculator:

Για τον υπολογισμό των συντεταγμένων του bbox χρησιμοποιήθηκε η εφαρμογή OSM Tile Calculator.

4.3 Overpass API, QL και OSM Tile Calculator



Εικόνα 4.4: Η εφαρμογή OSM Tile Calculator.

Με την OSM Tile Calculator μεταφέρουμε και τοποθετούμε το πορτοκαλί πλαίσιο, όπως φαίνεται στην Εικόνα 4.4, σε ορισμένο σημείο του χάρτη παίρνοντας τις αντίστοιχες συντεταγμένες από την καρτέλα CD.

Για την υλοποίησή μας, χρησιμοποιήσαμε την εφαρμογή python wrapper Overpass, η οποία συνδυάζει τη δυνατότητα να ζητάμε δεδομένα από την Overpass μέσω της γλώσσας προγραμματισμού python. Έτσι, ζητάμε συγκεκριμένες τιμές δεδομένων από την βάση δεδομένων του OpenStreetMap, με στόχο τη δημιουργία ενός δυναμικού εξαγωγέα προσαρμοσμένων δεδομένων.

Παρακάτω, βλέπουμε ένα παράδειγμα κώδικα, με τον οποίο ζητάμε δεδομένα από το Overpass API με αυτόν τον τρόπο.

Listing 4.1: Ερώτημα για γεωγραφικό πλάτος και γεωγραφικό μήκος για δεδομένα χρησιμοποιώντας τις συντεταγμένες bbox και την εφαρμογή Overpass.

```
import overpass
api = overpass.API(timeout=600,endpoint = Overpass_URL)
query = """
node[ 'Key'='Value '](bbox);
out center;
fmt = 'csv (::type , ::lat , ::lon) '
```

” ” ”

```
data = api.get(query, responseformat = fmt)
```

4.4 Εξαγωγή Προσαρμοσμένων Δεδομένων

Στην συνέχεια, παρουσιάζουμε βήμα-βήμα τη δημιουργία εργαλείου για την εξαγωγή προσαρμοσμένων δεδομένων με δυναμικό τρόπο, το οποίο χρησιμοποιήσαμε για να δημιουργήσουμε το σύνολο με το οποίο εκπαιδεύσαμε νευρωνικά δίκτυα κατά την υλοποίηση αυτής της εργασίας. Το εργαλείο αποτελείται από τρία κύρια κομμάτια, τον εξαγωγέα δεδομένων (Εικόνα 4.5), τον συνδυαστή (Εικόνα 4.8), καθώς και το τμήμα με το οποίο γίνεται η επικοινωνία με την εφαρμογή Maps Static .

Βασικές παράμετροι του Εξαγωγέα Δεδομένων:

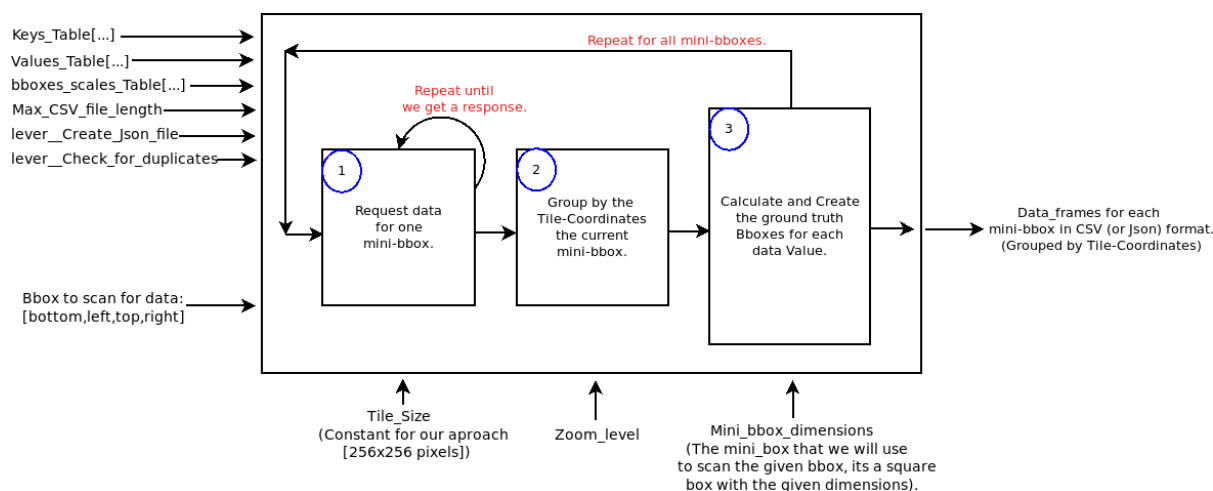
- *bbox_scales_Table*[]: Αποτελεί τον πίνακα που περιέχει τις κλίμακες με τις οποίες καθορίζουμε το μέγεθός του *ground_truth_bbox* σε σχέση με το μέγεθος του πλακιδίου, για κάθε ζεύγος κλειδιού (Key) και τιμής (Value). Αν για παράδειγμα, σε ένα κελί του έχει τιμή ίση με δύο (*bbox_scale_Table*[*i*] == 2, όπου *i* ακέραιος θετικός αριθμός), τότε το *ground_truth_bbox* του *i*-οστού ζεύγους θα έχει το μισό μήκος και πλάτος από το πλακίδιο της δορυφορικής εικόνας (1/4 του εμβαδού) στην οποία αντιστοιχεί. Στην περίπτωση που οι διαστάσεις του *ground_truth_bbox* ξεπερνούν τα όρια του πλακιδίου, τότε τα όρια του πλακιδίου ορίζονται ως όρια του *ground_truth_bbox*.
- *Values_Table*[], *Keys_Table*[]: Είναι πίνακες οι οποίοι περιέχουν τα ζεύγη των τιμών-κλειδιών τα οποία χρησιμοποιούνται για να γίνουν τα αιτήματα των δεδομένων από το OpenStreetMap. Τα στοιχεία του πίνακα *Values_Table*[] προσδιορίζουν μια γενικευμένη κατηγορία των στοιχείων της βάσης δεδομένων και εκείνα που ανήκουν στον *Keys_Table*[] στοχεύουν σε συγκεκριμένα δεδομένα της εκάστοτε κατηγορίας.
- *Max_CSV_file_length*: Με την μεταβλητή αυτή καθορίζεται το μέγιστο μέγεθος των CSV αρχείων που γίνονται αποδεκτά από το πρόγραμμά μας (Εξαγωγέας Δεδομένων) απεσταλμένα από την βάση δεδομένων του OSM. Στην περίπτωση που τα CSV αρχεία ξεπερνούν αυτό το οριακό μέγεθος, τότε αρχικά ανακατεύουμε τα δεδομένα που έχουν σταλθεί και έπειτα κρατάμε το πλήθος των δεδομένων που ορίζεται από την μεταβλητή *Max_CSV_file_length*.

4.4 Εξαγωγή Προσαρμοσμένων Δεδομένων

Ο λόγος ύπαρξης της μεταβλητής αυτής, καθορίζεται από την ανάγκη ομοιομορφίας του πλήθους των δεδομένων από όλα τα σημεία του ορισμένου χώρου επάνω στον χάρτη, από τον οποίο χρειαζόμαστε δεδομένα, καθώς και από το εκθετικό πρόβλημα της χρονικής πολυπλοκότητας που προκύπτει κατά την ομαδοποίηση (η οποία γίνεται βάσει των συντεταγμένων πλακιδίου) των CSV αρχείων ως προς το μέγεθός τους.

- *lever_CSV_files_to_check_for_errors*: Μεταβλητή με την οποία ο χρήστης επιλέγει αν το πρόγραμμα θα πραγματοποιήσει έλεγχο για τυχόν διπλότυπα δεδομένα, καθώς και αν θα αποδεσμευτεί η επιπλέον πληροφορία που αποστέλλεται από την βάση δεδομένων του OpenStreetMap.
- *lever_JSON_Creation_file* : Μεταβλητή με την οποία επιλέγουμε την δημιουργία αρχείου τύπου JSON ως έξοδο του Εξαγωγέα δεδομένων. Η δυνατότητα αυτή προστέθηκε, διότι αρκετές εφαρμογές της βαθιάς εκμάθησης χρησιμοποιούν αρχεία τύπου JSON.
- *bbox[bottom, left, top, right]*: Πίνακας που περιέχει τέσσερα σημεία του χάρτη, με τα οποία ορίζεται το πλαίσιο-κουτί (bbox), από το οποίο θέλουμε δεδομένα του OpenStreetMap.
- *Range_for_minning_bbox*: Με βάση την μεταβλητή αυτή δημιουργείται ένα μικρότερο πλαίσιο-κουτί (*mini_bbox*) το οποίο χρησιμοποιείται για την σάρωση του bbox. Το *mini_bbox* έχει διαστάσεις [*Range_for_minning_bbox* × *Range_for_minning_bbox*]. Ο λόγος δημιουργίας αυτού του πλαισίου-σάρωσης προέκυψε διότι η βάση δεδομένων του OpenStreetMap έχει ένα ορισμένο ανώτατο όριο πλήθους δεδομένων, τα οποία μπορεί να στείλει για κάθε αίτημα.
Έτσι, η διάσπαση του χώρου από τον οποίο χρειαζόμαστε δεδομένα επιλύει το πρόβλημα αυτό, χωρίς να έχουμε σημαντικές απώλειες στην πληροφορία κατά την ομαδοποίηση των δεδομένων ως προς τις συντεταγμένες πλακιδίων (απώλεια πληροφορίας έχουμε μόνο στα όρια των *mini_bbox*).
- *Tile_Size*: Οι διαστάσεις πλακιδίου της δορυφορικής εικόνας (256 × 256 εικονοστοιχεία για την δική μας προσέγγιση).
- *Zoom_Level*: Το επίπεδο μεγέθυνσης των πλακιδίων δορυφορικών εικόνων που κατεβάζουμε με την χρήση της εφαρμογής Map Static από την βάση δεδομένων της Google.

4. Η ΠΡΟΣΕΓΓΙΣΗ ΜΑΣ



Εικόνα 4.5: Εξαγωγή Δεδομένων.

Όπως απεικονίζεται και στο Σχήμα 4.5, κατά το πρώτο στάδιο λειτουργίας του Εξαγωγέα Δεδομένων στέλνεται αίτημα για δεδομένα από το OpenStreetMap με βάση ένα *mini.bbox*. Έτσι, σε αυτό το σημείο το πρόγραμμα αναμένει μέχρι τα δεδομένα που ζητήθηκαν είτε να σταλούν επιτυχώς, είτε σε περίπτωση που η βάση δεδομένων δεν ανταποκριθεί τότε να δοκιμάσει όλες τις εναλλακτικές εκδοχές των βάσεων που είναι προσβάσιμες στο δίκτυο.

A	B	C
@type	@lat	@lon
node	36.0152811	-114.7379244
node	36.0151959	-114.7379409
node	36.0150841	-114.7379793
node	36.0149364	-114.7381385
node	36.0146038	-114.7382681
node	36.0146833	-114.7384017
node	36.0145551	-114.7385515

@type	@lat	@lon	@Value	@bbox_scales
node	36.0152811	-114.7379244	hydro	3
node	36.0151959	-114.7379409	hydro	3
node	36.0150841	-114.7379793	hydro	3
node	36.0149364	-114.7381385	hydro	3
node	36.0146038	-114.7382681	hydro	3
node	36.0146833	-114.7384017	hydro	3
node	36.0145551	-114.7385515	hydro	3

node	36.2645225	-115.1852466	solar	2
node	36.2650016	-115.1865743	solar	2
node	35.1435791	-114.0649351	wind	4
node	35.1446505	-114.0664752	wind	4
node	35.1460582	-114.0675244	wind	4
node	35.1464624	-114.0694585	wind	4
node	35.1477966	-114.0710677	wind	4
node	35.1435791	-114.0649351	wind	4
node	35.1446505	-114.0664752	wind	4
node	35.1460582	-114.0675244	wind	4
node	35.1464624	-114.0694585	wind	4
node	35.1477966	-114.0710677	wind	4
node	36.0152811	-114.7379244	hydro	3
node	36.0151959	-114.7379409	hydro	3
node	36.0150841	-114.7379793	hydro	3

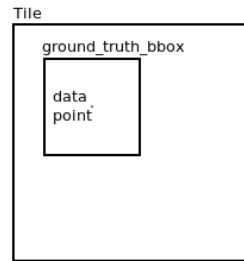
Εικόνα 4.6: Επεξεργασία δεδομένων κατά το πρώτο επίπεδο.

Τα δεδομένα που αποστέλλονται από το OpenStreetMap έχουν πληροφορία για το είδος τους, καθώς και για τις συντεταγμένες που βρίσκονται στον χάρτη. Έτσι, με βάση αυτές, υπολογίζει για κάθε δεδομένο τις συντεταγμένες του πλακιδίου στο οποίο ανήκει (@TileCoordinate). Στην συνέχεια, ομαδοποιούνται τα δεδομένα χρησιμοποιώντας τις συντεταγμένες πλακιδίου (Εικόνα 4.6) (για να αντιστοιχίσουμε τα δεδομένα που ανήκουν στο

4.4 Εξαγωγή Προσαρμοσμένων Δεδομένων

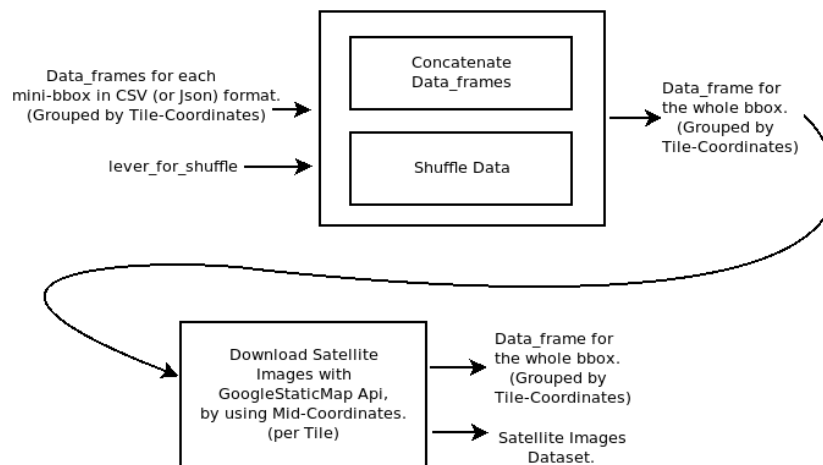
ίδιο πλαχίδιο). Ακόμη, το πρόγραμμα υπολογίζει τα πλαίσια μέσα στα οποία βρίσκονται τα δεδομένα (*ground_truth_bbox*) βάσει της αντίστοιχης μεταβλητής *bbox_scale* που έχουμε ορίσει (Εικόνα 4.7). Το *ground_truth_bbox* είναι ένας πίνακας τεσσάρων τιμών (*top_left_x, top_left_y, bottom_right_x - top_left_x, bottom_right_y - top_left_y*).

Οι πρώτες δύο τιμές ορίζουν το πάνω αριστερό εικονοστοιχείο του τετράγωνου-πλασιού, μέσα στο οποίο βρίσκεται το δεδομένο που αναφέρεται, και οι δύο τελευταίες το πλήθος των εικονοστοιχείων που πρέπει να μετακινηθούμε για να βρεθούμε στο κάτω δεξιό εικονοστοιχείο του πλασιού. Τέλος, η διαδικασία επαναλαμβάνεται για κάθε *mini_bbox*.



Εικόνα 4.7: Πλακάκι με ένα *ground_truth_bbox*.

Ως έξοδο του πρώτου μέρους του προγράμματος έχουμε πλαίσια δεδομένων για κάθε ένα από τα *mini_bbox*, σε CSV ή JSON μορφή (*Data_frames_for_each_mini_bbox*).



Εικόνα 4.8: Συνδυασμός δεδομένων και λήψη δορυφορικών εικόνων με χρήση των Mid-Coordinates.

Στο δεύτερο στάδιο του Εξαγωγέα Δεδομένων έχουμε ως είσοδο τα πλαίσια δεδομένων

$$[a_1, a_2, a_3, a_4, a_5] = 1, \quad [a_1, a_2, a_3, a_4] = 1, \quad [a_1, a_2, a_3] = 1, \quad [a_1, a_2] = 1, \quad [a_1] = 1,$$


1. *Journal of the American Medical Association*, 1997; 278: 1039-1044.

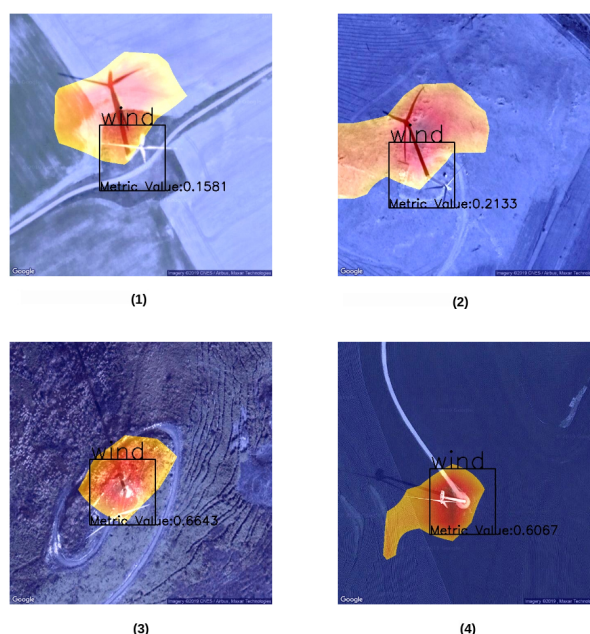
[illegible]

τις μεσαίες συντεταγμένες για να κατεβάσουμε μέσω της εφαρμογής Maps Static δορυφορικές εικόνες (Εικόνα 4.10) από την βάση δεδομένων της Google. Με αυτόν τον τρόπο δημιουργούμε το σύνολο εικόνων που θα χρησιμοποιήσουμε, σε συνδυασμό με το σύνολο δεδομένων πληροφοριών, στην εκπαίδευση νευρωνικών δικτύων.

4.5 Δημιουργία Μετρικής Fmatch

Υπάρχουν αρκετοί τρόποι για να εκτιμήσει κανείς αν η επιλογή ταξινομητή για την επίλυση ενός προβλήματος είναι ορθή, αφού αυτό το ζήτημα εξαρτάται από την φύση του ίδιου του προβλήματος.

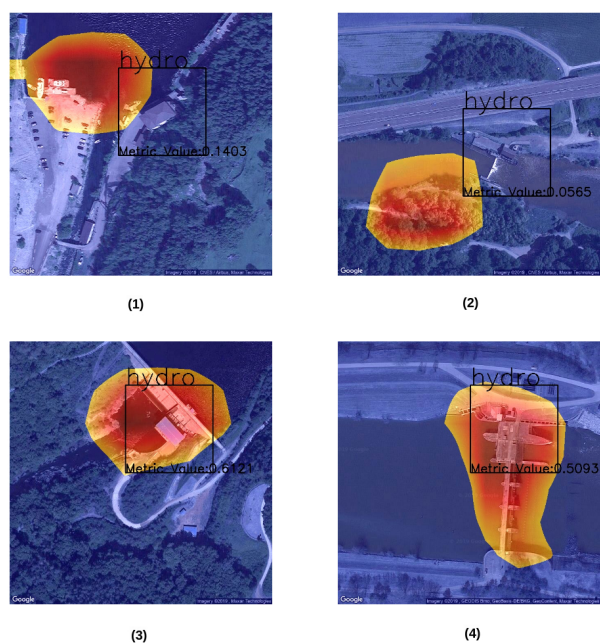
Στην υλοποίησή μας δημιουργήσαμε μία μετρική με βάση την οποία κρίνεται κατά πόσο ένας ταξινομητής καταφέρνει να ταξινομήσει μία κλάση, έχοντας την προσοχή του μόνο στα χαρακτηριστικά που αφορούν την ίδια την κλάση και όχι στις συσχετίσεις τις οποίες δημιουργεί ο ταξινομητής σε σχέση με το περιβάλλον της.



Εικόνα 4.11: Στις δύο πρώτες εικόνες που ο συσχετισμός έγινε με βάση την σκίαση, η μετρική έχει χαμηλότερη τιμή σε σύγκριση με τις επόμενες δύο.

4. Η ΠΡΟΣΕΓΓΙΣΗ ΜΑΣ

Η χρησιμότητα της μετρικής αυτής θα φανεί καλύτερα με το ακόλουθο παράδειγμα. Στην περίπτωση όπου έχουμε ένα αυτοκαθοδηγούμενο όχημα και θέλουμε να επιλέξουμε ανάμεσα σε δύο ταξινομητές, οι οποίοι έχουν την δυνατότητα να αναγνωρίζουν ανθρώπους, τότε η ανάγκη του να έχουμε μια αναγνώριση η οποία εστιάζει την προσοχή της στα χαρακτηριστικά ενός ατόμου (π.χ. άκρα, ρούχα, παπούτσια) και όχι για παράδειγμα στο πεζοδρόμιο, καθίσταται εξαιρετικά σημαντική. Τέλος, ακόμα ένας σημαντικός λόγος για την δημιουργία της μετρικής, είναι η ανάγκη για καλύτερη κατανόηση της λειτουργίας των ταξινομητών.



Εικόνα 4.12: Στις δύο πρώτες εικόνες που ο συσχετισμός έγινε βασισμένος στο συνδυασμό νερού και δάσους, η μετρική έχει χαμηλότερη τιμή σε σύγκριση με τις δύο πρώτες.

Όπως απεικονίζεται και στις Εικόνες 4.11 και 4.12, η τιμή της μετρικής προκύπτει από τον λόγο της τομής ανάμεσα στο *ground_truth_bbox* του αντικειμένου και τον θερμικό χάρτη που έχουμε από τον ταξινομητή για το εκάστοτε αντικείμενο, ως προς την ένωσή τους:

$$F_{match} = \frac{\text{τομή του πλήθους pixels του θερμικού χάρτη και του ground_truth_bbox}}{\text{ένωση του πλήθους pixels του θερμικού χάρτη και του ground_truth_bbox}}$$

Σημείωση: Τα εικονοστοιχεία που περιέχονται στο *ground_truth_bbox* έχουν βαρύτητα ίση με την μονάδα, έτσι η τομή προκύπτει από τις τιμές του θερμικού χάρτη που βρίσκονται μέσα στο *ground_truth_bbox* και η ένωση από το άθροισμα των τιμών του *ground_truth_bbox* με εκείνες του θερμικού χάρτη που είναι εκτός αυτού.

Το σύνολο τιμών της μετρικής είναι ανάμεσα στο μηδέν και την μονάδα ($0 \leq metric \leq 1$). Έτσι για τιμές κοντά στην μονάδα, ο ταξινομητής έχει συσχετίσει την ταξινόμηση μόνο με χαρακτηριστικά του ίδιου του αντικειμένου. Ενώ για τιμές κοντά στο μηδέν η συσχέτιση έχει γίνει με χαρακτηριστικά που ανήκουν στο περιβάλλον του αντικειμένου.

Με βάση τα παραπάνω φαίνεται πως στην περίπτωση της επιλογής ταξινομητή για αυτοκαθοδηγούμενα οχήματα, ανάμεσα σε δύο ταξινομητές με σχετικά ίδια ευστοχία αναγνώρισης, η επιλογή μπορεί να γίνει βασισμένη στην εν λόγω μετρική.

Όπως είδαμε στην υλοποίησή μας, υπολογίσαμε την παραπάνω μετρική με βάση το *ground_truth_bbox* του αντικειμένου που ταξινομήσαμε. Αυτό προέκυψε από τον τρόπο με τον οποίο δημιουργήσαμε το σύνολο των δεδομένων μας.

Παρόλο που αυτός ο τρόπος υπολογισμού είναι μια αρκετά καλή προσέγγιση, θα ήταν ακόμα καλύτερο να χρησιμοποιηθεί η κατάτμηση του αντικειμένου που ταξινομείται, διότι με τον τρόπο αυτό, το *ground_truth_bbox* ή καλύτερα το *ground_truth_segmentation*, θα πάρει μορφή όμοια με εκείνης του αντικειμένου που θέλουμε να ταξινομήσουμε. Έτσι, η τομή του θερμικού χάρτη και του *ground_truth_segmentation* αποτελείται μόνο από εικονοστοιχεία τα οποία συμπίπτουν με χαρακτηριστικά του αντικειμένου

4.6 Εκπαίδευση με Προσαρμοσμένα Σύνολα Δεδομένων

Με την χρήση του εργαλείου εξαγωγής προσαρμοσμένων δεδομένων δημιουργήσαμε, για την υλοποίηση αυτής της διπλωματικής εργασίας, δύο σύνολα δεδομένων έτσι ώστε να χρησιμοποιηθούν στην εκπαίδευση νευρωνικών δικτύων για χρήση σε ταξινόμηση πολλαπλών κλάσεων.

Το πρώτο αποτελείται από έξι χιλιάδες δορυφορικές εικόνες (6000) τριών τύπων (αιολική ενέργεια, ηλιακή ενέργεια, υδροηλεκτρική ενέργεια), όπου αντιστοιχούν δύο χιλιάδες εικόνες (2000) σε κάθε τύπο. Το δεύτερο αποτελείται από είκοσι τρεις χιλιάδες οκταχόσιες δορυφορικές εικόνες (23800) επτά τύπων (νεκροταφείο, εμπορικό, χωράφι, αγρόκτημα, δάσος, λιβάδι, γρασίδι), όπου αντιστοιχούν τρεις χιλιάδες τετραχόσιες (3400) εικόνες για κάθε τύπο. Μαζί με τα σύνολα δεδομένων των εικόνων έχουν δημιουργηθεί και τα αντίστοιχα CSV αρχεία, που περιέχουν την πληροφορία που θα χρησιμοποιηθεί στην συνέχεια της υλοποίησης μας (βλέπε Ενότητα 4.4).

Αυτά τα δύο σύνολα δεδομένων χρησιμοποιήθηκαν στην εκπαίδευση δύο γνωστών νευρωνικών δικτύων, του VGG16 και του ResNet50. Οι αρχιτεκτονικές των δικτύων που χρησιμοποιήθηκαν φαίνονται στις παρακάτω εικόνες.

Η αρχιτεκτονική του VGG16 που χρησιμοποιήθηκε:

Όπως απεικονίζεται στην αρχιτεκτονική του δικτύου VGG16 (Εικόνα 4.13) (δίκτυο που εκπαιδεύσαμε για την ταξινόμηση των ΑΠΕ), έχουμε αφαιρέσει το τελευταίο επίπεδο και στην θέση του έχουμε προσθέσει δύο επιπλέον επίπεδα. Ένα ολικής μέσης συγκέντρωσης, το οποίο είναι απαραίτητο για την εξαγωγή των θερμικών χαρτών, καθώς και ένα επίπεδο εξόδου, το οποίο αντιστοιχεί τις πιθανότητες σε κάθε έναν από τους τρεις τύπους.

Επιλέξαμε να εκπαιδεύσουμε μια προεκπαιδευμένη εκδοχή του VGG16, διότι στα πρώτα επίπεδα το δίκτυο έχει μάθει απλές συσχετίσεις οι οποίες μπορούν να χρησιμοποιηθούν και στην δική μας υλοποίηση.

Κατά την εκπαίδευση του, επιλέξαμε να είναι προς εκπαίδευση μόνο οι παράμετροι των μπλοκ τρία και πέντε του δικτύου, έτσι ώστε να γίνει η προσαρμογή στην δική μας περίπτωση. Όπως φαίνεται, το δίκτυο έχει 14.716.227 συνολικές παραμέτρους εκ των οποίων οι 8.556.291 είναι προς εκπαίδευση.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 128, 128, 3)	0
block1_conv1 (Conv2D)	(None, 128, 128, 64)	1792
block1_conv2 (Conv2D)	(None, 128, 128, 64)	36928
block1_pool (MaxPooling2D)	(None, 64, 64, 64)	0
block2_conv1 (Conv2D)	(None, 64, 64, 128)	73856
block2_conv2 (Conv2D)	(None, 64, 64, 128)	147584
block2_pool (MaxPooling2D)	(None, 32, 32, 128)	0
block3_conv1 (Conv2D)	(None, 32, 32, 256)	295168
block3_conv2 (Conv2D)	(None, 32, 32, 256)	590080
block3_conv3 (Conv2D)	(None, 32, 32, 256)	590080
block3_pool (MaxPooling2D)	(None, 16, 16, 256)	0
block4_conv1 (Conv2D)	(None, 16, 16, 512)	1180160
block4_conv2 (Conv2D)	(None, 16, 16, 512)	2359808
block4_conv3 (Conv2D)	(None, 16, 16, 512)	2359808
block4_pool (MaxPooling2D)	(None, 8, 8, 512)	0
block5_conv1 (Conv2D)	(None, 8, 8, 512)	2359808
block5_conv2 (Conv2D)	(None, 8, 8, 512)	2359808
block5_conv3 (Conv2D)	(None, 8, 8, 512)	2359808
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 512)	0
dense_1 (Dense)	(None, 3)	1539
Total params: 14,716,227		
Trainable params: 8,556,291		
Non-trainable params: 6,159,936		

Εικόνα 4.13: VGG16 με προσθήκη επιπέδου ολικής μέσης συγκέντρωσης για το σύνολο δεδομένων γεννητριών ΑΠΕ.

```
# compile model
opt=Adam(lr=0.001, beta_1=0.9, beta_2=0.999, epsilon=None, decay=0.0, amsgrad=False)
model.compile(optimizer=opt, loss='binary_crossentropy', metrics=[fbeta])
# checkpoint
filepath="weights.epoch:{epoch:02d},val_loss:-{val_loss:.2f},val_fbeta:-{val_fbeta:.2f}.hdf5"
checkpoint = ModelCheckpoint(filepath, monitor='val_fbeta', verbose=1, save_best_only=True, mode='max')
```

Εικόνα 4.14: Υπερπαράμετροι για το VGG16.

4. Η ΠΡΟΣΕΓΓΙΣΗ ΜΑΣ

Για την εκπαίδευσή του χωρίσαμε τυχαία το σύνολο δεδομένων των εικόνων ΑΠΕ σε δύο τμήματα. Το πρώτο με 5000 εικόνες που χρησιμοποιήθηκε στην εκπαίδευση και την επικύρωση του δικτύου. Το δεύτερο με 1000 εικόνες που χρησιμοποιήθηκε για τον έλεγχο του σε εικόνες που δεν έχει συναντήσει κατά την εκπαίδευση. Σημείωση: Κάθε υποσύνολο έχει ίσο πλήθος αντικειμένων κάθε κλάσης.

Ως συνάρτηση απώλειας επιλέχθηκε η binary-crossentropy και ως βελτιστοποιητής ο αλγόριθμος Adam με υπερ-παραμέτρους, όπως φαίνεται στο Σχήμα 4.14. Η επιλογή των τιμών των υπέρ-παραμέτρων του βελτιστοποιητή έγινε με βάση τις τιμές που προτείνονται από τον δημιουργό του, για την καλύτερη σύγκλιση του μοντέλου και ο δείκτης μάθησης επιλέχθηκε από δοκιμές που υλοποιήσαμε.

Για την αποθήκευση των καλύτερων εκδόσεων του μοντέλου έχουμε επιλέξει την παρακολούθηση της μετρικής f-score (Ενότητα 2.3).

```
batch_size = 50
my_training_batch_generator = My_Custom_Generator_ForTrain(trainX, trainY, batch_size)
my_validation_batch_generator = My_Custom_Generator_ForTest(testX, testY, batch_size)
# define model
model.VGG_callbacks_list = define_model()
# fit model
#history
history = model.fit_generator(generator=my_training_batch_generator,
                             steps_per_epoch = int((trainX.shape[0]+testX.shape[0])*0.8 / batch_size),
                             epochs = 200,
                             callbacks=VGG_callbacks_list,
                             verbose = 1,
                             validation_data = my_validation_batch_generator,
                             validation_steps = int((trainX.shape[0]+testX.shape[0])*0.2 / batch_size))
# evaluate model
loss, fbeta = model.evaluate_generator(my_validation_batch_generator, steps=len(my_validation_batch_generator), verbose=0)
```

Εικόνα 4.15: Προσαρμογή και αξιολόγηση του VGG16.

Από τις 5000 εικόνες, το 80% χρησιμοποιήθηκε για την εκπαίδευση του μοντέλου και το 20% για την επικύρωση του. Ο διαχωρισμός έγινε τυχαία.

Αρχικά, οι δορυφορικές εικόνες κανονικοποιήθηκαν, έτσι ώστε τα εικονοστοιχεία τους σε κάθε κανάλι να έχουν μηδενικό μέσο όρο τιμών και διακύμανση ίση με την μονάδα. Για τις εικόνες που χρησιμοποιήθηκαν κατά την εκπαίδευση, δημιουργήσαμε αντίγραφα εικόνων κάθετων και οριζόντιων αναστροφών, τα οποία προστέθηκαν στο σύνολο δεδομένων για την επέκταση του συνόλου εκπαίδευσης. Ακόμη, όπως φαίνεται και στην Εικόνα ?? το μέγεθος της παρτίδας εισόδων (*batch_size*) έχει τιμή 50, τόσο για την εκπαίδευση όσο και για την επικύρωση του δικτύου. Τέλος, το πλήθος των εποχών (*epochs*) κατά την διάρκεια της εκπαίδευσης επιλέχθηκε ίσο με 200.

Η αρχιτεκτονική του ResNet50 που χρησιμοποιήθηκε:

Ομοίως με το δίκτυο VGG16, στην εφαρμογή του ResNet50 έχουμε αφαιρέσει το τελευταίο επίπεδο του δικτύου και στην θέση του έχουμε προσθέσει ένα επίπεδο ολικής μέσης συγκέντρωσης για την εξαγωγή των θερμικών χαρτών, καθώς και ένα επίπεδο εξόδου, (Εικόνα 4.16).

Και στην περίπτωση αυτή, επιλέξαμε μια προ-εκπαιδευμένη εκδοχή του δικτύου ResNet50 για τους ίδιους λόγους με την υλοποίηση του VGG16.

Κατά την εκπαίδευση του δικτύου επιλέχθηκαν να είναι προς εκπαίδευση όλες οι παράμετροι του νευρωνικού, για να εκμεταλλευτούμε την ιδιότητα της απομνημόνευσης κατά την εκμάθηση των residual blocks.

add_14 (Add)	(None, 4, 4, 2048)	0	bn5a_branch2c[0][0] bn5a_branch1[0][0]
activation_43 (Activation)	(None, 4, 4, 2048)	0	add_14[0][0]
res5b_branch2a (Conv2D)	(None, 4, 4, 512)	1049088	activation_43[0][0]
bn5b_branch2a (BatchNormalizati	(None, 4, 4, 512)	2048	res5b_branch2a[0][0]
activation_44 (Activation)	(None, 4, 4, 512)	0	bn5b_branch2a[0][0]
res5b_branch2b (Conv2D)	(None, 4, 4, 512)	2359808	activation_44[0][0]
bn5b_branch2b (BatchNormalizati	(None, 4, 4, 512)	2048	res5b_branch2b[0][0]
activation_45 (Activation)	(None, 4, 4, 512)	0	bn5b_branch2b[0][0]
res5b_branch2c (Conv2D)	(None, 4, 4, 2048)	1050624	activation_45[0][0]
bn5b_branch2c (BatchNormalizati	(None, 4, 4, 2048)	8192	res5b_branch2c[0][0]
add_15 (Add)	(None, 4, 4, 2048)	0	bn5b_branch2c[0][0] activation_43[0][0]
activation_46 (Activation)	(None, 4, 4, 2048)	0	add_15[0][0]
res5c_branch2a (Conv2D)	(None, 4, 4, 512)	1049088	activation_46[0][0]
bn5c_branch2a (BatchNormalizati	(None, 4, 4, 512)	2048	res5c_branch2a[0][0]
activation_47 (Activation)	(None, 4, 4, 512)	0	bn5c_branch2a[0][0]
res5c_branch2b (Conv2D)	(None, 4, 4, 512)	2359808	activation_47[0][0]
bn5c_branch2b (BatchNormalizati	(None, 4, 4, 512)	2048	res5c_branch2b[0][0]
activation_48 (Activation)	(None, 4, 4, 512)	0	bn5c_branch2b[0][0]
res5c_branch2c (Conv2D)	(None, 4, 4, 2048)	1050624	activation_48[0][0]
bn5c_branch2c (BatchNormalizati	(None, 4, 4, 2048)	8192	res5c_branch2c[0][0]
add_16 (Add)	(None, 4, 4, 2048)	0	bn5c_branch2c[0][0] activation_46[0][0]
activation_49 (Activation)	(None, 4, 4, 2048)	0	add_16[0][0]
global_average_pooling2d_1 (Glo	(None, 2048)	0	activation_49[0][0]
dense_1 (Dense)	(None, 3)	6147	global_average_pooling2d_1[0][0]
=====			
Total params: 23,593,859			
Trainable params: 23,540,739			
Non-trainable params: 53,120			

Εικόνα 4.16: ResNet50 με προσθήκη επιπέδου ολικής μέσης συγκέντρωσης για το σύνολο δεδομένων γεννητριών ΑΠΕ.

Όπως φαίνεται, το δίκτυο αποτελείται από 23.593.859 παραμέτρους εκ των οποίων οι 23.540.739 είναι προς εκπαίδευση (53.120 σταθερές).

Η διαδικασία που ακολουθήσαμε για τον διαχωρισμό του συνόλου δεδομένων των δορυφορικών εικόνων ΑΠΕ, καθώς και οι επιλογές για βελτιστοποιητή, συνάρτηση απώλειας και υπερ-παραμέτρους του μοντέλου επιλέχθηκαν ίδιες με εκείνες στην περίπτωση του VGG16.

Στο επόμενο κεφάλαιο παρουσιάζονται τα αποτελέσματα που προέκυψαν από εκπαίδευση των δικτύων και βάσει αυτών γίνεται επεξήγηση και σύγκριση της λειτουργίας τους.

4. Η ΠΡΟΣΕΓΓΙΣΗ ΜΑΣ

Κεφάλαιο 5

Αποτελέσματα

5.1 Απόδοση των Αρχιτεκτονικών VGG16, ResNet50

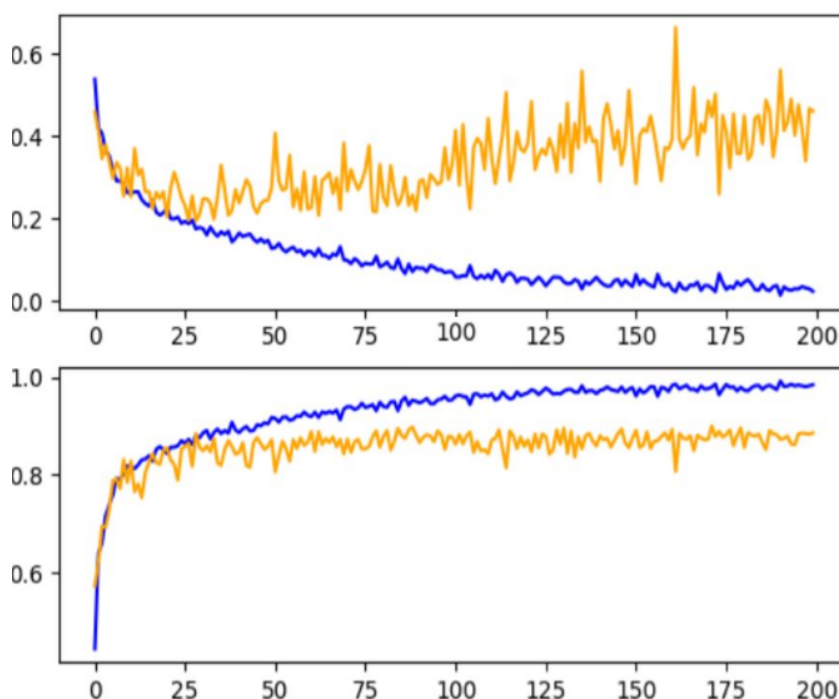
Στις ακόλουθες υπό-ενότητες παρουσιάζονται τα αποτελέσματα αυτής της εργασίας, τα οποία προέκυψαν κατά την εκπαίδευση των δύο γνωστών νευρωνικών δικτύων (VGG16 και ResNet50) σε σύνολο δεδομένων που δημιουργήσαμε. Συγκεκριμένα, παρουσιάζονται για κάθε ένα μοντέλο ξεχωριστά:

- Τα διαγράμματα των συναρτήσεων απώλειας και της μετρικής Fbeta.
- Η ευστοχία και οι πινάκες σύγχυσης των κλάσεων.
- Οι τιμές που αντιστοιχούν στην προτεινόμενη μετρική Fmatch.

5. ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1.1 Απόδοση του VGG16 στην Ταξινόμηση ΑΠΕ.

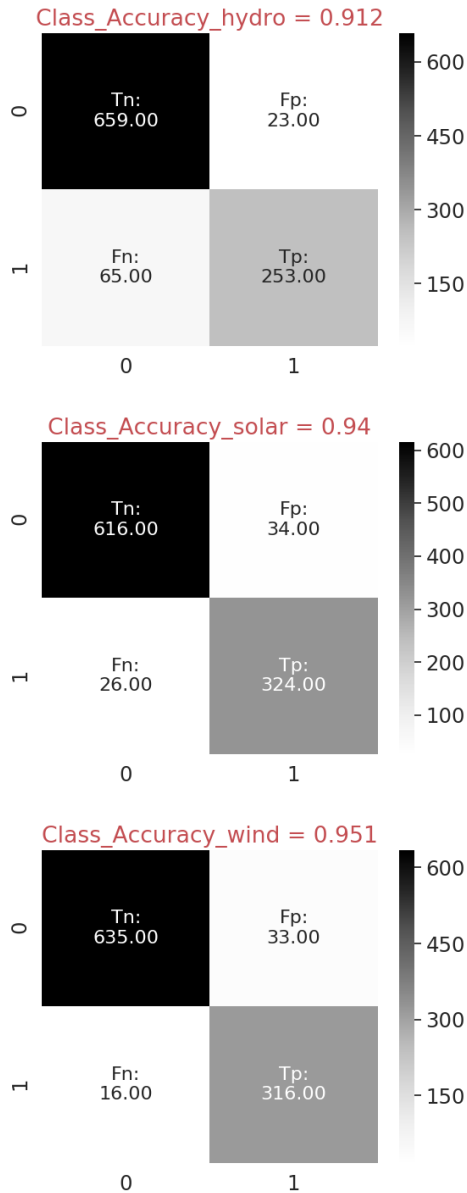
Από το ακόλουθο διάγραμμα που προέκυψε κατά την εκπαίδευση του νευρωνικού δικτύου (Εικόνα 5.1), συμπεραίνουμε πως παρόλο που η απώλεια μειώνεται διαρκώς (όπως και ήταν αναμενόμενο), κατά την επικύρωση ακολουθεί την μείωση αυτή μέχρι την εποχή 30 και στην συνέχεια αυξάνεται. Αυτό συμβαίνει διότι σε αυτό το σημείο έχουμε overfitting στα δεδομένα του συνόλου εκπαίδευσης. Ακόμη βλέπουμε πως η μετρική Fbeta έχει καλές τιμές και πριν την εποχή 30.



Εικόνα 5.1: Απώλεια Cross Entropy - Μετρική Fbeta.

Ο λόγος που κρατήθηκε η εκπαίδευση για τόσες εποχές είναι για να γίνει επεξήγηση της λειτουργίας του μοντέλου ακόμα και σε περιπτώσεις overfitting.

5.1 Απόδοση των Αρχιτεκτονικών VGG16, ResNet50



Εικόνα 5.2: Πίνακες σύγχυσης για το VGG16.

Ως καλύτερη εκδοχή των βαρών του νευρωνικού δικτύου, επιλέχθηκαν εκείνα που προέκυψαν κατά την 85η εποχή της εκπαίδευσης. Η απώλεια επικύρωσης και η μετρική Fbeta στην συγκεκριμένη περίπτωση έχουν τιμές 0.32 και 0.9 αντίστοιχα.

Με βάση τις 1000 δορυφορικές εικόνες που αποτελούν το τεστ σύνολο δεδομένων, η παραπάνω έκδοση του μοντέλου χρησιμοποιήθηκε στον υπολογισμό της ακρίβειας και της τιμής της μετρικής που παρουσιάσαμε στην Ενότητα 4.5 καθώς και στην δημιουργία του πίνακα σύγχυσης κάθε κλάσης.

Όπως βλέπουμε, το μοντέλο μας πετυχαίνει αρκετά καλές τιμές ακρίβειας για σύνολο δεδομένων στο οποίο δεν έχει εκπαιδευτεί, με μέγιστη εκείνη της κλάσης *wind* ($\simeq 0.95$).

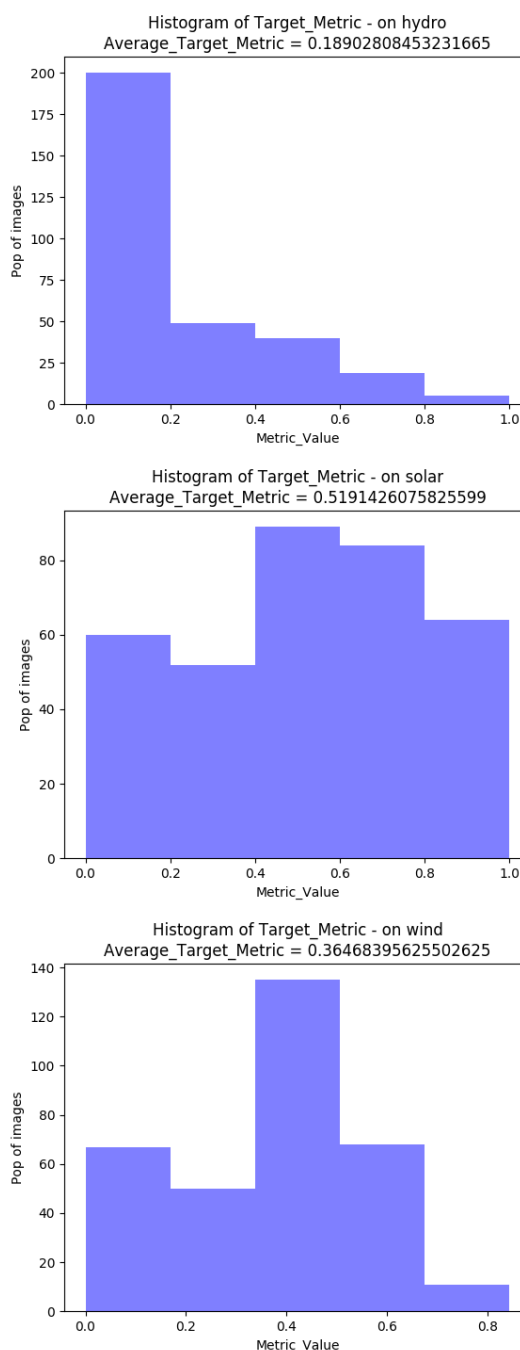
Μια πιθανή αιτία για αυτές τις υψηλές αποδόσεις, αποτελεί το γεγονός του ότι όλες οι δορυφορικές εικόνες, οι οποίες συσχετίζονται με κάποια από αυτές τις κλάσεις, είναι πανομοιότυπες. Αυτό συμβαίνει διότι, οι εγκαταστάσεις γεννητριών ΑΠΕ τοποθετούνται σε όμοια φυσικά περιβάλλοντα. Έτσι, το τεστ σύνολο των δορυφορικών εικόνων, μοιάζει αρκετά με εκείνο του συνόλου εκπαίδευσης.

5. ΑΠΟΤΕΛΕΣΜΑΤΑ

Στην συνέχεια, παρουσιάζονται τα ιστογράμματα για κάθε κλάση ως προς τη μετρική που προτείναμε, στην Εικόνα 5.3.

Παρατηρούμε πως μόνο για την κλάση *solar* το μοντέλο έχει δημιουργήσει καλή συσχέτιση της ταξινόμησης της κλάσης, με βάση τα χαρακτηριστικά των αντικειμένων που ανήκουν σε αυτήν, πετυχαίνοντας τιμή μετρικής $\simeq 0.52$. Για τις κλάσεις *hydro* και *wind*, παρόλο που η ακρίβεια των κλάσεων είναι αρκετά υψηλή, η τιμή της μετρικής είναι χαμηλή ($\simeq 0.189$, $\simeq 0.365$ αντίστοιχα). Έτσι, συμπεραίνουμε ότι οι συσχετίσεις που γίνονται κατά την ταξινόμηση αυτών των κλάσεων, συνδέονται κυρίως με το περιβάλλον των αντικειμένων.

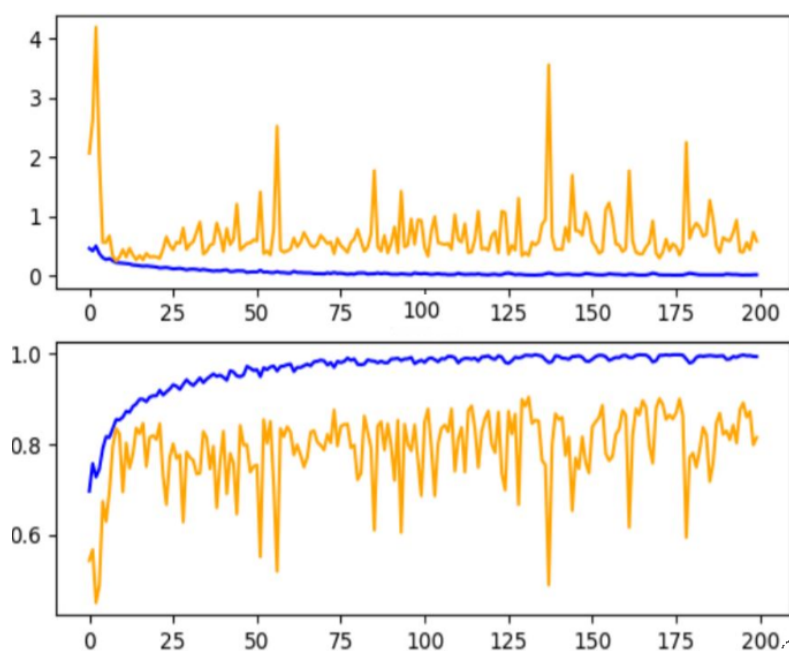
Τα συμπεράσματα που προκύπτουν από τις παραπάνω τιμές είναι πολλαπλά. Για την κλάση *hydro* ήταν αναμενόμενο να έχουμε χαμηλή τιμή μετρικής, αφού κάθε αντικείμενο της κλάσης βρίσκεται σε περιοχές που υπάρχει νερό. Έτσι, ο ταξινομητής συσχετίζει την ύπαρξη νερού με την ταξινόμηση του αντικειμένου. Αυτό που δεν ήταν προφανές είναι η χαμηλή τιμή μετρικής που προέκυψε για την κλάση *wind*. Όπως θα δούμε, ο ταξινομητής έδωσε βάση κυρίως στην σχίαση που δημιουργούν οι ανεμογεννήτριες για την ταξινόμησή τους.



Εικόνα 5.3: Ιστογράμματα Μετρικής για το δίκτυο VGG16.

5.1.2 Απόδοση του ResNet50 στην Ταξινόμηση ΑΠΕ.

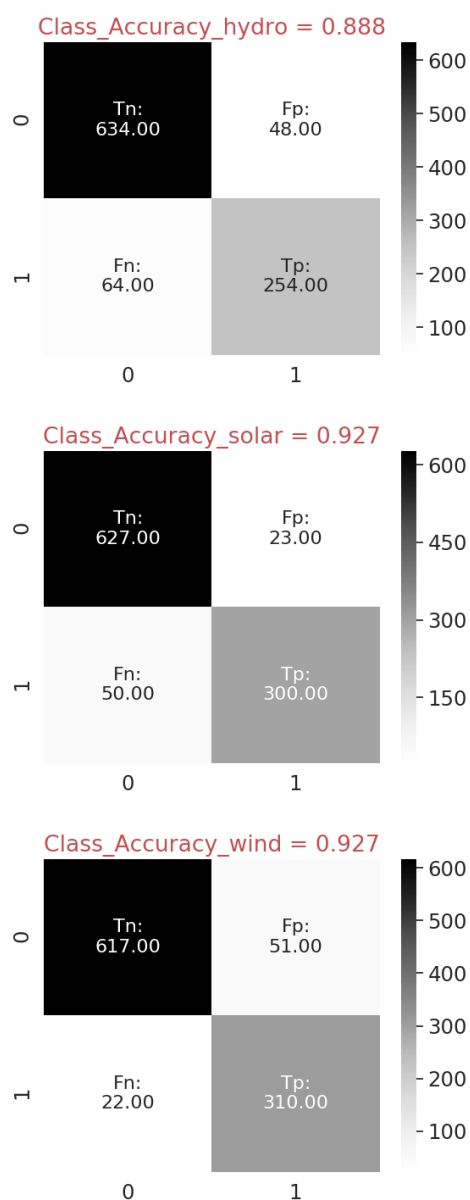
Από το ακόλουθο διάγραμμα, το οποίο προέκυψε κατά την εκπαίδευση του δικτύου ResNet50 (Εικόνα 5.4), συμπεραίνουμε ότι σε αντίθεση με το δίκτυο VGG16, το ResNet50 πετυχαίνει τις καλύτερες τιμές ως προς την απώλεια επικύρωσης και την μετρική Fbeta σε μεγαλύτερη εποχή (να σημειωθεί ότι αν και οι δύο εποχές είναι κοντά, η χρονική διάρκεια της εκπαίδευσης του ResNet50 είναι πολύ μεγαλύτερη εκείνης του VGG16). Ακόμη, παρατηρούμε πως δεν έχουμε φαινόμενο overfitting. Αυτό συμβαίνει λόγω της αρχιτεκτονικής των υπολειπόμενων μπλοκ του δικτύου που επιτρέπουν στο δίκτυο να χρησιμοποιεί μόνο εκείνα τα μπλοκ τα οποία έχουν επιδρούν θετικά στην ταξινόμηση.



Εικόνα 5.4: Απώλεια Cross Entropy - Μετρική Fbeta

Ο λόγος που κρατήθηκε η εκπαίδευση για τόσες εποχές είναι για να γίνει επεξήγηση της λειτουργίας του μοντέλου ακόμα και σε εκδοχές δικτύου που πλησιάζουν σε φαινόμενα overfitting.

5. ΑΠΟΤΕΛΕΣΜΑΤΑ



Εικόνα 5.5: Πίνακες σύγχυσης για το Resnet50.

Ως καλύτερη εκδοχή των βαρών του νευρωνικού δικτύου επιλέχθηκαν εκείνα που προέκυψαν κατά την 128η εποχή της εκπαίδευσης. Η απώλεια επικύρωσης και η μετρική Fbeta στην συγκεκριμένη περίπτωση έχουν τιμές 0.33 και 0.88 αντίστοιχα.

Με βάση τις 1000 δορυφορικές εικόνες που αποτελούν το τεστ σύνολο δεδομένων, η παραπάνω έκδοση του μοντέλου χρησιμοποιήθηκε στον υπολογισμό της ακρίβειας και της τιμής της μετρικής που παρουσιάσαμε στην Ενότητα 4.5 καθώς και στην δημιουργία των πινάκων σύγχυσης της κάθε κλάσης.

Αν και το μοντέλο πετυχαίνει καλές τιμές ακρίβειας για σύνολο δεδομένων στο οποίο δεν έχει εκπαιδευτεί, με υψηλότερες εκείνες των κλάσεων *wind* και *solar* ($\simeq 0.93$), είναι οριακά χαμηλότερες σε σύγκριση με εκείνες του VGG16.

Ένας πιθανός λόγος για αυτήν την οριακή διαφορά, βρίσκεται στο μεγαλύτερο πλήθος των επιπέδων που έχουμε ορίσει προς εκπαίδευση στο ResNet50 σε σχέση με εκείνων του VGG16. Το γεγονός αυτό ωθεί το ResNet50 να αποκλίνει περισσότερο από το βέλτιστο σημείο σύγκλισης, για εποχές επανάληψης κοντά στο βέλτιστο σημείο σύγκλισης, σε σχέση με το VGG16.

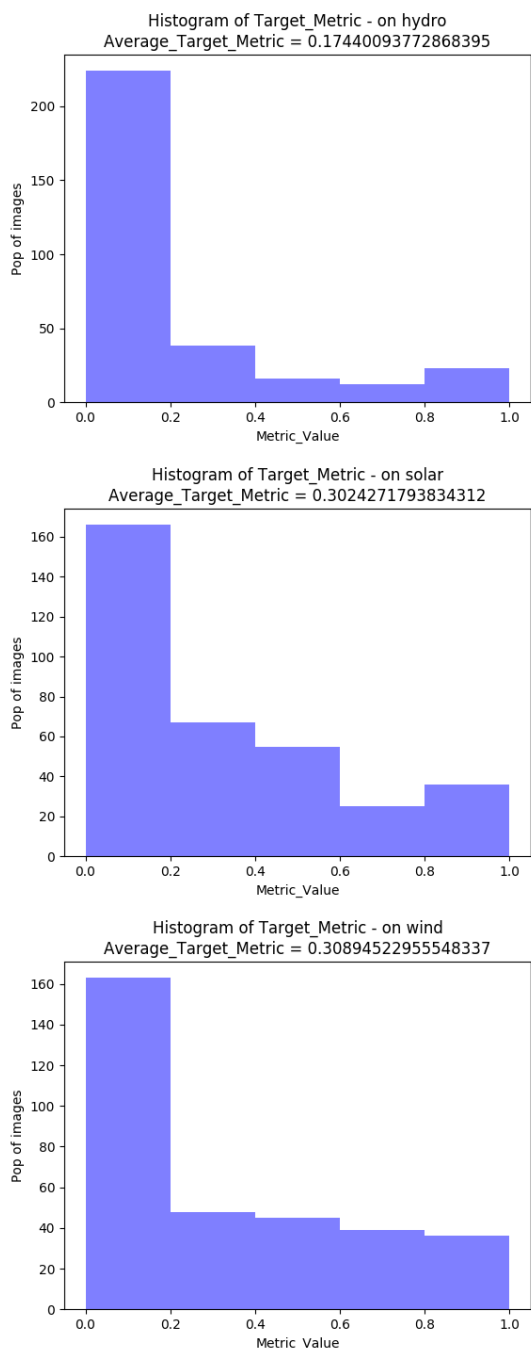
5.1 Απόδοση των Αρχιτεκτονικών VGG16, ResNet50

Στην συνέχεια, παρουσιάζονται τα ιστογράμματα για κάθε κλάση ως προς τη μετρική που προτείνουμε, στην Εικόνα 5.6.

Όπως και στην περίπτωση του VGG16, το ResNet50 πετυχαίνει την χαμηλότερη τιμή μετρικής για την κλάση *hydro* ($\simeq 0.174$). Έτσι, το μοντέλο συσχετίζει την ταξινόμηση των αντικειμένων αυτής της κλάσης με την ύπαρξη νερού στο περιβάλλον.

Παρατηρούμε επίσης ότι για τις άλλες δύο κλάσεις (*wind* και *solar*) ο ταξινομητής πετυχαίνει σχετικά χαμηλές τιμές μετρικής ($\simeq 0.31$). Έτσι συμπεραίνουμε πως έχει την τάση να συσχετίζει τις ταξινομήσεις δίνοντας μεγαλύτερη βάση στα χαρακτηριστικά του περιβάλλοντος, σε σχέση με το μοντέλο που προέκυψε από την αρχιτεκτονική VGG16.

Τέλος, βλέπουμε ότι έχουμε μικρότερη διασπορά των τιμών του ResNet50 σε σχέση με εκείνων του VGG16. Ο λόγος που συμβαίνει αυτό το φαινόμενο είναι διότι κατά την εκπαίδευση του τα υπολειπόμενα μπλοκ επιτρέπουν στην συνάρτηση εξόδου να κινηθεί είτε προς την κατεύθυνση που μας ενδιαφέρει, είτε να παραμείνει στην υπάρχουσα θέση. Έτσι, το δίκτυο μαθαίνει να ταξινομεί με βάση λιγότερα χαρακτηριστικά των αντικειμένων που ταξινομούνται σε σχέση με το VGG16.



Εικόνα 5.6: Ιστογράμματα Μετρικής για το δίκτυο ResNet50.

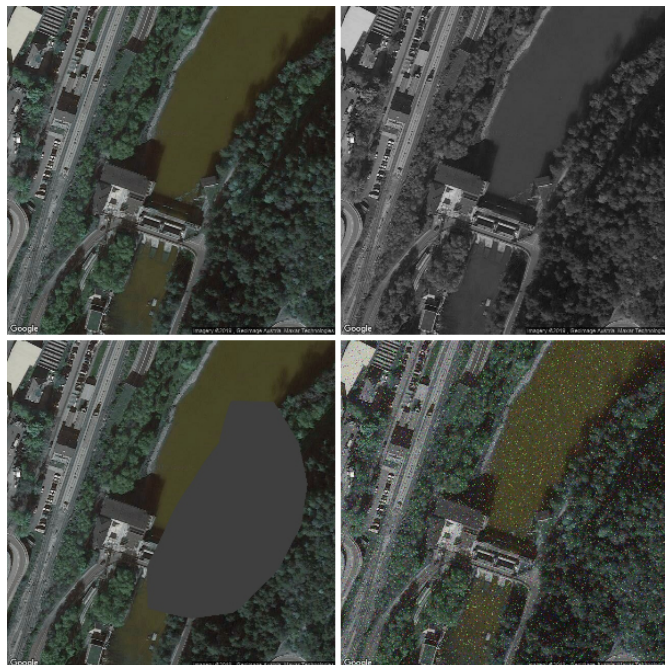
5. ΑΠΟΤΕΛΕΣΜΑΤΑ

5.2 Επεξήγηση Δικτύων με Χρήση Θερμικών Χαρτών

Με την χρήση της βιβλιοθήκης OpenCV δημιουργήσαμε παραλλαγές των εικόνων εισόδου, όπως φαίνεται στην Εικόνα 5.7, τις οποίες επανατοποθετήσαμε στα δίκτυα έτσι ώστε να εξάγουμε περαιτέρω συμπεράσματα για τον τρόπο που αυτά ταξινομούν.

Υλοποιήθηκαν τρεις παραλλαγές για κάθε εικόνα:

- Μια ασπρόμαυρη έκδοση της αρχικής εικόνας.
- Μια έκδοσή της, έχοντας καλύψει με τον μέσο όρο των τιμών των εικονοστοιχείων την επιφάνεια που αποτελούσε τον θερμικό χάρτη της αρχικής εικόνας.
- Μια έκδοσή της, εφαρμόζοντας θόρυβο αλάτι και πιπέρι.



Εικόνα 5.7: Υδροηλεκτρική γεννήτρια.

Στην συνέχεια παρουσιάζουμε τα συμπεράσματα τα οποία εξάγουμε από πραγματικά θετικά (TP) παραδείγματα βάσει των θερμικών χαρτών που προέκυψαν.

5.2 Επεξήγηση Δικτύων με Χρήση Θερμικών Χαρτών

Εξαγωγή θερμικών χαρτών από επαναπροσδιορισμένες εικόνες - hydro:

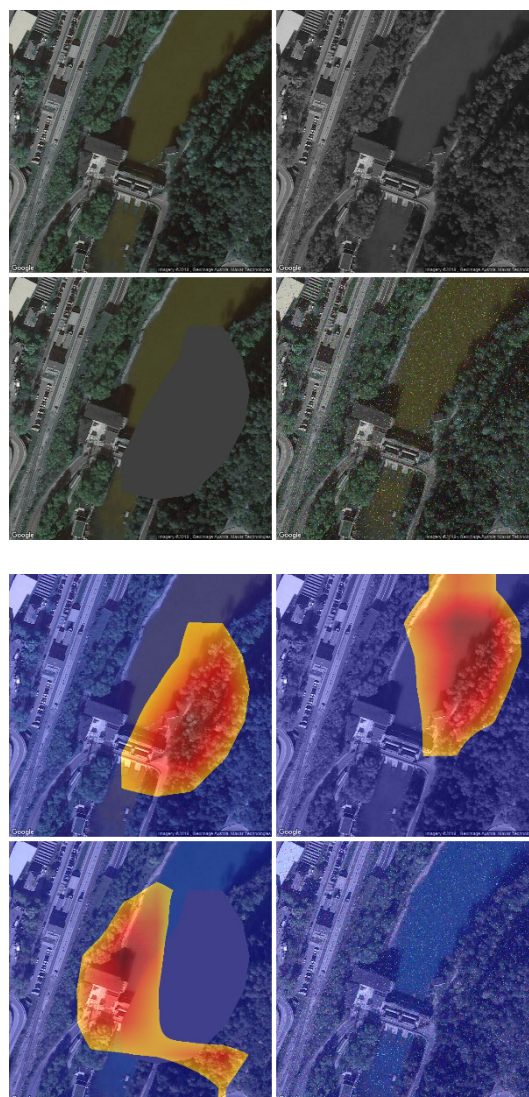
Όπως απεικονίζεται, στην Εικόνα 5.8, έχουμε εξάγει τους θερμικούς χάρτες που αντιστοιχούν στις τροποποιημένες εικόνες για το μοντέλο VGG16, καθώς και τις πιθανότητες για κάθε περίπτωση.

Αρχικά, παρατηρούμε πως το μοντέλο έχει ταξινομήσει την εικόνα ως *hydro* με πιθανότητα $\simeq 0.85$, έχοντας εστιάσει στον συνδυασμό νερού, δάσους και κτηρίου.

Έπειτα για την ασπρόμαυρη εικόνα, πάλι το μοντέλο προβλέπει με υψηλή πιθανότητα την κλάση *hydro* $\simeq 0.76$ ανεβάζοντας όμως και την πιθανότητα της κλάσης *wind* $\simeq 0.32$. Αυτό μας δείχνει πως το VGG16 δεν έχει δημιουργήσει έντονες συσχετίσεις των χρωμάτων για την κλάση *hydro* και ότι η αύξηση της κλάσης *wind* πιθανόν να συσχετίζεται με το ότι μπερδεύει το ασπρόμαυρο με την ύπαρξη των σκιών που έχει μάθει να ταξινομεί.

Στην τρίτη περίπτωση, όπου από το μοντέλο κρύβουμε τον αρχικό θερμικό χάρτη, παρατηρούμε πως καταφέρνει να κάνει ορθή ταξινόμηση (*hydro* $\simeq 0.76$), διότι υπάρχει αρκετή πληροφορία που συσχετίζεται με την κλάση.

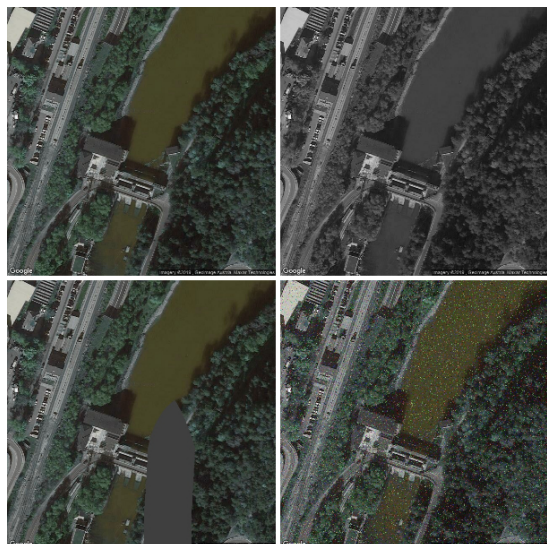
Με την εισαγωγή του θορύβου στην εικόνα δεν καταφέρνει το μοντέλο να κάνει σωστή ταξινόμηση του αντικειμένου, αυξάνοντας την πιθανότητα ύπαρξης αντικειμένου της κλάσης *solar* $\simeq 0.56$. Βλέπουμε λοιπόν πως η αλλοίωση των σχημάτων επηρεάζουν αρκετά την απόφαση



```
predictions for True_Image:
hydro:0.84545 solar:0.02545 wind:0.00125
predictions for Fake_Image:
hydro:0.75648 solar:0.01548 wind:0.01475
predictions for Fake_Image_BLACK:
hydro:0.83478. solar:0.01273 wind:0.32307
predictions for Fake_Salt_and_Pepper_Image:
hydro:0.00478 solar:0.56096 wind:0.00014
```

Εικόνα 5.8: θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης *hydro* - VGG16.

5. ΑΠΟΤΕΛΕΣΜΑΤΑ



```
predictions for True_Image:
  hydro:0.81243 solar:0.00128 wind:0.01511
predictions for Fake_Image:
  hydro:0.80021 solar:0.00314 wind:0.06101
predictions for Fake_Image_BLACK:
  hydro:0.30341 solar:0.01456 wind:0.68025
predictions for Fake_Salt_and_Pepper_Image:
  hydro:0.78094 solar:0.02285 wind:0.00726
```

Εικόνα 5.9: Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης *hydro* - ResNet50.

του μοντέλου και για την κλάση *solar* φαίνεται πως τα επαναλαμβανόμενα μοτίβα αυξάνουν την πιθανότητα ύπαρξης αντικειμένου.

Οι θερμικοί χάρτες που αντιστοιχούν στις εικόνες που τροποποιήθηκαν για το μοντέλο ResNet50 και οι πιθανότητες της ταξινόμησης των κλάσεων παρουσιάζονται στην Εικόνα 5.9.

Παρατηρούμε πως και το ResNet50 ταξινομεί την κλάση *hydro* δίνοντας έμφαση στον συνδυασμό κτηρίου και δάσους.

Για την ασπρόμαυρη εικόνα, το ResNet50 προβλέπει με χαμηλή τιμή την ύπαρξη αντικειμένου της κλάσης *hydro* $\simeq 0.3$, ενώ ταυτόχρονα αυξάνει την πιθανότητα της κλάσης *wind* $\simeq 0.68$. Συμπεραίνουμε λοιπόν πως το ResNet50 έχει δώσει έμφαση στον συνδυασμό των χρωμάτων για την ταξινόμηση της κλάσης *hydro*.

Μια ιδιόζουσα συμπεριφορά του μοντέλου παρατηρείται στην τρίτη εικόνα, όπου κρύβουμε εκείνο το τμήμα του αρχικού θερμικού χάρτη. Παρατηρούμε πως το ResNet50 συνεχίζει να εστιάζει την προσοχή του στο ίδιο σημείο έχοντας μεγάλη πιθανότητα για την κλάση *hydro*. Έτσι βλέπουμε πως το νευρωνικό δίνει έμφαση βάση του μέσου όρου των υποπεριοχών της εικόνας.

Τέλος, η προσθήκη θορύβου δεν φαίνεται να επηρεάζει σημαντικά την ταξινόμηση, πετυχαίνοντας για την κλάση *hydro* $\simeq 0.78$ καθώς οι πιθανότητες των άλλων δύο κλάσεων αυξάνονται ελάχιστα.

5.2 Επεξήγηση Δικτύων με Χρήση Θερμικών Χαρτών

Εξαγωγή θερμικών χαρτών από επαναπροσδιορισμένες εικόνες - wind:

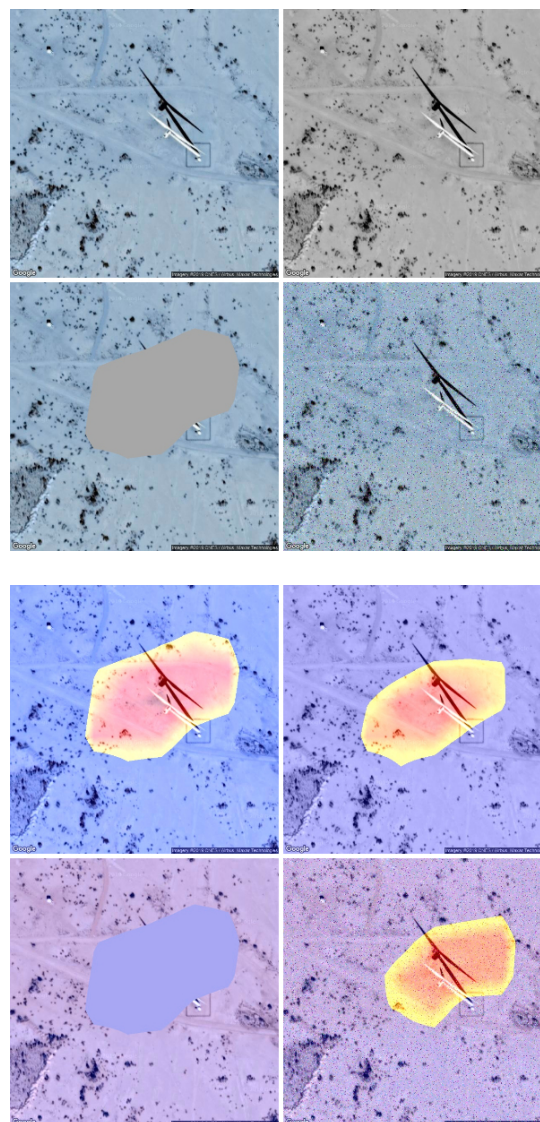
Στην 5.10 έχουμε τους θερμικούς χάρτες που αντιστοιχούν σε τροποποιημένες εικόνες από αντικείμενο της κλάσης *wind* για το μοντέλο VGG16.

Αρχικά, παρατηρούμε πως το μοντέλο έχει ταξινομήσει την εικόνα ως *wind* με πιθανότητα $\simeq 0.92$, δίνοντας προσοχή στον ανεμόμυλο και στην σκιά που δημιουργεί.

Για την ασπρόμαυρη εικόνα, το μοντέλο προβλέπει με μεγαλύτερη πιθανότητα την κλάση *wind $\simeq 0.95$, σε σχέση με εκείνη της αρχικής εικόνας εστιάζοντας περισσότερο στην σκιά που δημιουργεί ο ανεμόμυλος.*

Στην συνέχεια όπου κρύβουμε τον αρχικό θερμικό χάρτη, βλέπουμε ότι το VGG16 δεν καταφέρνει να κάνει ορθή ταξινόμηση διότι χάνεται σχεδόν όλη η πληροφορία που συσχετίζεται με την κλάση *wind*. Έτσι, το μοντέλο πιθανόν να περνάει την υπόλοιπη εικόνα για νερό, αυξάνοντας την πιθανότητα της κλάσης *hydro* $\simeq 0.6$.

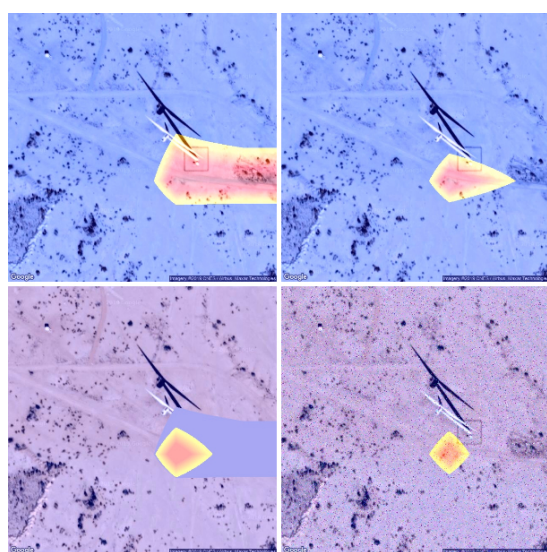
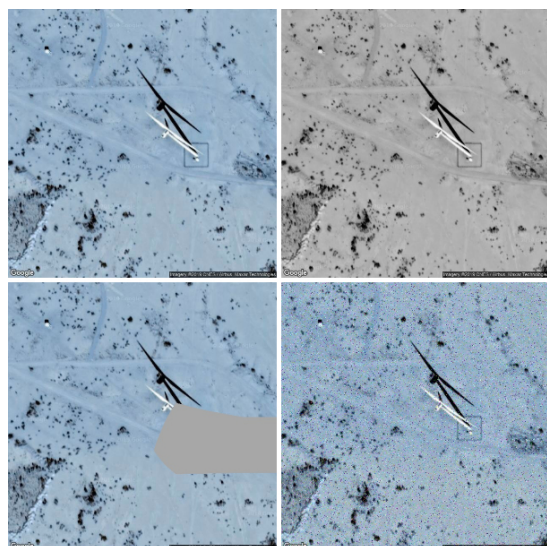
Στην τέταρτη εικόνα όπου έχουμε θόρυβο, η ταξινόμηση του μοντέλου δεν επηρεάζεται σημαντικά, πετυχαίνοντας πιθανότητα της κλάσης *wind* $\simeq 0.89$ και έχοντας μια ασήμαντη αύξηση της πιθανότητας της κλάσης *solar* από $\simeq 0.047$ (αρχική εικόνα) σε $\simeq 0.73$. Βλέπουμε λοιπόν πως η αλλοίωση των εικόνων εισόδου δεν επηρεάζει σημαντικά την απόφαση του μοντέλου για την κλάση *wind*.



```
predictions for True_Image:
hydro:0.03608 solar:0.04684 wind:0.92407
predictions for Fake_Image:
hydro:0.60977 solar:0.03044 wind:0.01757
predictions for Fake_Image_BLACK:
hydro:0.00043 solar:0.00036 wind:0.94872
predictions for Fake_Salt_and_Pepper_Image:
hydro:0.01451 solar:0.07293 wind:0.89235
```

Εικόνα 5.10: Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης *wind* - VGG16.

5. ΑΠΟΤΕΛΕΣΜΑΤΑ



```
predictions for True_Image:
  hydro:0.00011 solar:0.05141 wind:0.86343
predictions for Fake_Image:
  hydro:0.06124 solar:0.07824 wind:0.81279
predictions for Fake_Image_BLACK:
  hydro:0.00392 solar:0.02133 wind:0.83144
predictions for Fake_Salt_and_Pepper_Image:
  hydro:0.00115 solar:0.59996 wind:0.12402
```

Εικόνα 5.11: Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης *wind* - ResNet50.

Οι θερμικοί χάρτες που αντιστοιχούν στις εικόνες που τροποποιήθηκαν για το μοντέλο ResNet50 παρουσιάζονται στην Εικόνα 5.11.

Από τον θερμικό χάρτη της αρχικής εικόνας βλέπουμε πως το ResNet50, σε αντίθεση με το VGG16, δεν ταξινομεί την κλάση *wind* βάσει της σκίασης που δημιουργεί ο ανεμόμυλος.

Για την ασπρόμαυρη εικόνα το ResNet50 προβλέπει με χαμηλότερη πιθανότητα από εκείνη της αρχικής εικόνας την ύπαρξη αντικειμένου της κλάσης *wind* $\simeq 0.83$. Έτσι βλέπουμε πάλι πως το ResNet50 δεν έχει δώσει έμφαση στην σκιά όπως το VGG16.

Όπως και στην κλάση *hydro*, το ResNet50 στην τρίτη εικόνα, που κρύβουμε εκείνο το τμήμα του αρχικού θερμικού χάρτη, συνεχίζει να εστιάζει την προσοχή του στο ίδιο σημείο έχοντας μεγάλη πιθανότητα για την κλάση *wind*. Έτσι βλέπουμε ξανά πως το νευρωνικό δίκτυο δίνει έμφαση στον μέσο ορο των υποπεριοχών της εικόνας.

Στην τελευταία περίπτωση, αν και η εστίαση του θερμικού χάρτη δεν αλλάζει για την κλάση *wind*, η προσθήκη του θορύβου επηρεάζει σημαντικά την ταξινόμηση πετυχαίνοντας για την κλάση *wind* $\simeq 0.12$. Ακόμη, μπερδεύει το μοντέλο αυξάνοντας την πιθανότητα της κλάσης *solar* από $\simeq 0.05$ σε $\simeq 0.6$.

5.2 Επεξήγηση Δικτύων με Χρήση Θερμικών Χαρτών

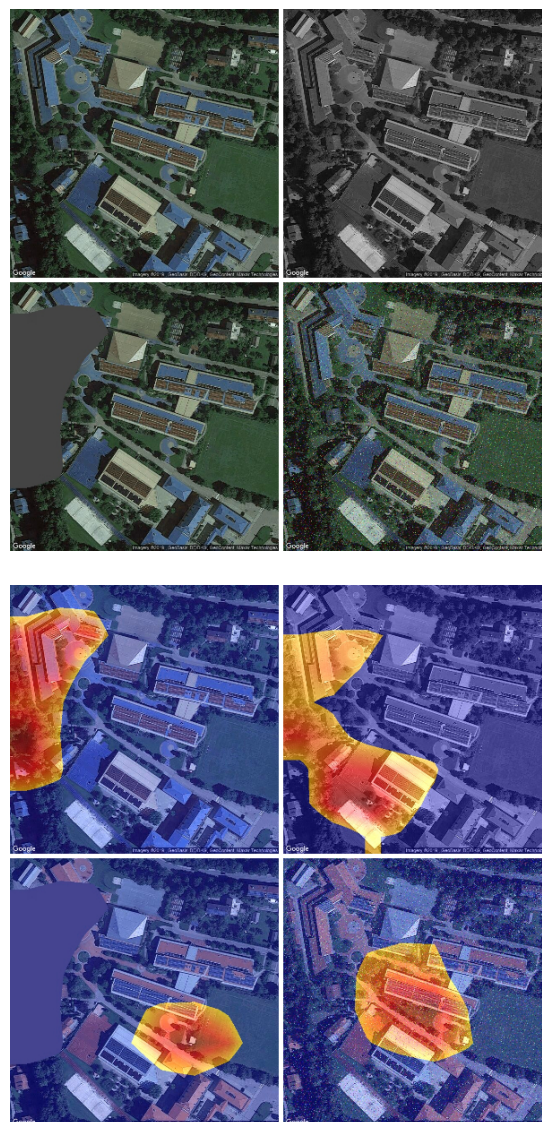
Εξαγωγή θερμικών χαρτών από επαναπροσδιορισμένες εικόνες - solar:

Για την κλάση *solar* (Εικόνα 5.12), ο θερμικός χάρτης που εξάγεται από το VGG16 για την αρχική εικόνα εστιάζει στον συνδυασμό του δάσους και των κτηρίων. Έτσι, βλέπουμε πως παρόλο που το νευρωνικό δίκτυο έχει εντοπίσει αντικείμενο της κλάσης αυτής με πιθανότητα $\simeq 0.73$, εντοπίζει λανθασμένα και την ύπαρξη αντικειμένου της κλάσης *hydro* με πιθανότητα $\simeq 0.38$.

Στην ασπρόμαυρη εικόνα για να ταξινομήσει το δίκτυο, απλώνει τον θερμικό χάρτη έχοντας ως αποτέλεσμα μεγαλύτερη τιμή για την πιθανότητα αντικειμένου της κλάσης *solar* και μικρότερη εκείνης της *hydro*. Έτσι, συμπεραίνεται πως το VGG16 ξεχωρίζει τα δάση κυρίως από το χρώμα τους.

Ενδιαφέρον παρουσιάζει η λειτουργία του δικτύου στην τρίτη περίπτωση, όπου έχουμε καλύψει την εικόνα στο σημείο του αρχικού θερμικού χάρτη. Βλέπουμε λοιπόν πως το νευρωνικό εστιάζει την προσοχή του στο αντικείμενο της κλάσης *solar*, πετυχαίνοντας υψηλή πιθανότητα ύπαρξης του αντικειμένου $\simeq 0.9$.

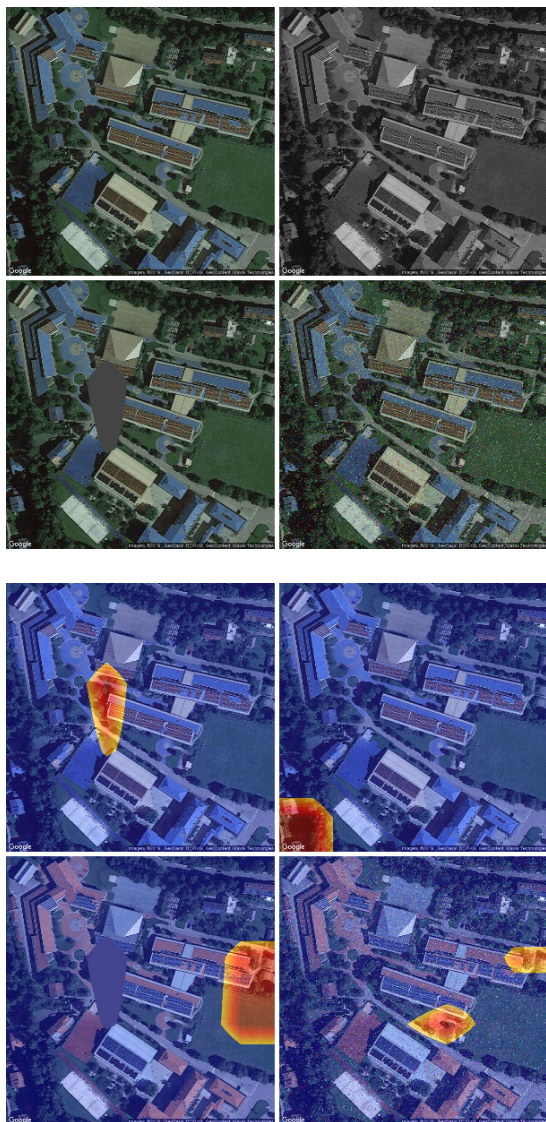
Έτσι, αν θέλαμε να εκπαιδεύσουμε το δίκτυο έτσι ώστε να συσχετίζει την ταξινόμηση της *solar* μόνο με χαρακτηριστικά αντικειμένου της κλάσης αυτής, μπορούμε να αφαιρέσουμε όλες τις δασικές περιοχές από τις εικόνες του αρχικού συνόλου δεδομένων και να εκπαιδεύσουμε ξανά το δίκτυο.



```
predictions for True_Image:
hydro:0.38146 solar:0.73184 wind:0.00048
predictions for Fake_Image:
hydro:0.33168 solar:0.90202 wind:0.00026
predictions for Fake_Image_BLACK:
hydro:0.14708 solar:0.82638 wind:0.00383
predictions for Fake_Salt_and_Pepper_Image:
hydro:0.00585 solar:0.94194 wind:0.00004
```

Εικόνα 5.12: Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης solar - VGG16.

5. ΑΠΟΤΕΛΕΣΜΑΤΑ



```
predictions for True_Image:
hydro:0.41025 solar:0.57892 wind:0.00064
predictions for Fake_Image:
hydro:0.61208 solar:0.30001 wind:0.00125
predictions for Fake_Image_BLACK:
hydro:0.00075 solar:0.00051 wind:0.87894
predictions for Fake_Salt_and_Pepper_Image:
hydro:0.18585 solar:0.78215 wind:0.00038
```

Εικόνα 5.13: Θερμικοί χάρτες από επαναπροσδιορισμένες εικόνες της κλάσης solar - ResNet50.

Σε ότι αφορά την κλάση *solar* για το ResNet50 παρατηρούμε από την Εικόνα 5.13 όμοια συμπεριφορά με εκείνη του VGG16 για την αρχική εικόνα. Αν και, όπως φαίνεται από τον θερμικό χάρτη, εστιάζει λιγότερο σε δασική περιοχή.

Ακόμη, για την ασπρόμαυρη εκδοχή της εικόνας η πιθανότητα που αντιστοιχεί στην κλάση *wind* είναι εκείνη που κυριαρχεί με τιμή $\simeq 0.88$. Αυτό συμβαίνει διότι το δίκτυο εστιάζει σε μαύρη περιοχή της εικόνας, θεωρώντας πως διακρίνει σκιά κάποιου ανεμόμυλου.

Όμως, στην περίπτωση που καλύπτουμε την περιοχή της εικόνας, το νευρωνικό δίκτυο ταξινομεί λανθασμένα, έχοντας μεγαλύτερη πιθανότητα για την ύπαρξη αντικείμενου της κλάσης *hydro* $\simeq 0.61$.

Τέλος, με την εισαγωγή θορύβου στην εικόνα, το ResNet50 συμπεριφέρεται όμοια με το VGG16, αυξάνοντας την εκτίμηση του για την ύπαρξη της κλάσης *solar*.

Κεφάλαιο 6

Συμπεράσματα

6.1 Συμπεράσματα

Σε αυτήν την διπλωματική παρουσιάσαμε ένα προγραμματιστικό εργαλείο, το οποίο παρέχει την δυνατότητα εξαγωγής σύνολου δεδομένων δορυφορικών εικόνων με δυναμικό τρόπο για την εκπαίδευση νευρωνικών δικτύων, χρησιμοποιώντας πληροφορίες από την βάση δεδομένων του OpenStreetMap. Με το εργαλείο αυτό δημιουργήσαμε ένα σύνολο δεδομένων με αντικείμενα τριών κλάσεων. Με το εν λόγω σύνολο εκπαιδεύσαμε δυο γνωστά νευρωνικά δίκτυα, το VGG16 και το ResNet50. Εξάγοντας τους θερμικούς χάρτες των δύο αυτών δικτύων, ερμηνεύσαμε τον τρόπο με τον οποίο ταξινομούν καθώς και συγκρίναμε την λειτουργία μεταξύ αυτών. Ακόμη, δημιουργήσαμε μία μετρική με την οποία υποδεικνύεται αν ένα δίκτυο ταξινομεί τα αντικείμενα μιας κλάσης, δίνοντας προσοχή στα χαρακτηριστικά που αποτελούν το ίδιο το αντικείμενο.

Είναι γεγονός πως όλο και περισσότεροι ερευνητές τα τελευταία χρόνια ενσωματώνουν στον τομέα της έρευνάς τους βαθιά νευρωνικά δίκτυα. Όμως, ακόμα και οι καλύτεροι αυτών παραδέχονται πως ο τρόπος με τον οποίο τα νευρωνικά συμπεριφέρονται, σε κάθε στάδιο της αρχιτεκτονικής τους, είναι αρκετά πολύπλοκος. Έτσι, ο σκοπός της εργασίας αφορούσε την μελέτη και την διερεύνηση των νευρωνικών δικτύων ως προς τον τρόπο με τον οποίο εκπαιδεύονται, αλλά και ως προς την διαδικασία που ακολουθούν κατά την ταξινόμηση.

Δείξαμε λοιπόν πως η δυνατότητα της ανάλυσης και της μεταποίησης του συνόλου δεδομένων, το οποίο χρησιμοποιείται στην εκπαίδευση ενός δικτύου καθώς και η χρήση μετρικών και θερμικών χαρτών για την επεξήγηση των αποτελεσμάτων του, είναι μια στρατηγική που πρέπει να ακολουθήσουμε για την κατανόηση του τρόπου λειτουργίας του.

6.2 Μελλοντική Δουλειά

6.2.1 Διερεύνηση Περισσότερων Αρχιτεκτονικών

Στην προσέγγιση που ακολουθήσαμε, χρησιμοποιήθηκαν οι αρχιτεκτονικές των δικτύων VGG16 και ResNet50. Αν και αυτά τα δύο νευρωνικά χρησιμοποιούνται κατά κόρον σε εφαρμογές τεχνητής νοημοσύνης, η διερεύνηση περισσότερων και νεώτερων αρχιτεκτονικών έχει να αποφέρει σημαντικά ωφέλη για την κατανόηση της λειτουργίας τους. Έτσι, η εργασία μας εύκολα μπορεί να επεκταθεί ως προς αυτήν την κατεύθυνση, ακολουθώντας την ίδια διαδικασία των βημάτων που υλοποιήθηκαν για περισσότερες αρχιτεκτονικές γνωστών νευρωνικών δικτύων.

6.2.2 Συνδυάζοντας Όλα τα Στάδια

Όπως αναλύθηκε, η εργασία αυτή αποτελείται από τρία μεμονωμένα στάδια. Πρώτο στάδιο, αποτελεί την συλλογή και την επεξεργασία πληροφοριών από την βάση δεδομένων του OpenStreetMap. Δεύτερο, την δημιουργία του συνόλου με αντιστοίχιση δορυφορικών εικόνων και των εν λόγω πληροφοριών. Το τρίτο και τελευταίο στάδιο αποτελείται από την εκπαίδευση και την κατανόηση γνωστών νευρωνικών δικτύων. Ως μελλοντικό στόχο ορίζουμε την ένωση των τριών αυτών τμημάτων σε ένα συνολικό προγραμματιστικό εργαλείο, με το οποίο ο χειριστής του θα έχει την δυνατότητα να εξάγει εύκολα τα συμπεράσματα που αφορούν πληθώρα γνωστών αρχιτεκτονικών δικτύων, έτσι ώστε η επιλογή και η εξέλιξη αυτών να καταστεί ευκολότερη.

Βιβλιογραφία

- [1] Thoma, M.: Latex-examples/tikz/vgg-16. <https://github.com/MartinThoma/LaTeX-examples/tree/master/tikz/vgg-16> 20.03.2017. xiii, 24
- [2] Erhan, D., Bengio, Y., Courville, A., Vincent, P.: Visualizing higher-layer features of a deep network. University of Montreal **1341**(3) (2009) 1 xiv, 37
- [3] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., eds.: Computer Vision – ECCV 2014, Cham, Springer International Publishing (2014) 818–833 xiv, xiv, 2, 38, 39
- [4] Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., Batra, D.: Grad-cam: Why did you say that? arXiv preprint arXiv:1611.07450 (2016) xiv, xiv, 40, 41
- [5] Ribeiro, M.T., Singh, S., Guestrin, C.: ”why should i trust you?” explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. (2016) 1135–1144 xiv, xiv, 41, 42
- [6] Samek, W., Wiegand, T., Müller, K.R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. ITU Journal: ICT Discoveries - Special Issue 1 - The Impact of Artificial Intelligence (AI) on Communication Networks and Services **1** (10 2017) 1–10 2
- [7] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [8] Chattopadhyay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N.: Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE (2018) 839–847
- [9] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. (2017) 618–626
- [10] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11) (1998) 2278–2324
- [11] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2016)
- [12] Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. In Buntine, W., Grobelnik, M., Mladenić, D., Shawe-Taylor, J., eds.: *Machine Learning and Knowledge Discovery in Databases*, Berlin, Heidelberg, Springer Berlin Heidelberg (2009) 254–269
- [13] Pastor-Pellicer, J., Zamora-Martínez, F., España-Boquera, S., Castro-Bleda, M.J.: F-measure as the error function to train neural networks. In Rojas, I., Joya, G., Gabestany, J., eds.: *Advances in Computational Intelligence*, Berlin, Heidelberg, Springer Berlin Heidelberg (2013) 376–384
- [14] Rizwan, M., Nadeem, A., Sindhu, M.A.: Analyses of classifier’s performance measures used in software fault prediction studies. *IEEE Access* **7** (2019) 82764–82775
- [15] Kingma, D., Ba, J.: Adam: A method for stochastic optimization. *International Conference on Learning Representations* (12 2014)
- [16] Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* **5**(2) (1994) 157–166

- [17] Derczynski, L.: Complementarity, f-score, and nlp evaluation. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). (2016) 261–266
- [18] Cakir, F., He, K., Xia, X., Kulis, B., Sclaroff, S.: Deep metric learning to rank. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (June 2019)
- [19] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. *Neural computation* **1**(4) (1989) 541–551
- [20] Le Cun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Handwritten digit recognition with a back-propagation network. In: Proceedings of the 2nd International Conference on Neural Information Processing Systems. (1989) 396–404
- [21] LeCun, Y., et al.: Generalization and network design strategies. *Connectionism in perspective* **19** (1989) 143–155
- [22] Hoseinzade, E., Haratizadeh, S.: Cnnpred: Cnn-based stock market prediction using a diverse set of variables. *Expert Systems with Applications* **129** (2019) 273–285
- [23] Lin, M., Chen, Q., Yan, S.: Network in network. *arXiv preprint arXiv:1312.4400* (2013)
- [24] Karalas, K., Tsagkatakis, G., Zervakis, M., Tsakalides, P.: Deep learning for multi-label land cover classification. In Bruzzone, L., ed.: *Image and Signal Processing for Remote Sensing XXI*. Volume 9643., International Society for Optics and Photonics, SPIE (2015) 244 – 257