



**ΧΑΡΟΚΟΠΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ**

**ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ**

Κατηγοριοποίηση Εικόνων με Χρήση Τεχνικών Deep Learning

Πτυχιακή εργασία

Παναγιώτης Κωνσταντίνος Γέμος



Αθήνα, 2017



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

**ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ**

Τριμελής Εξεταστική Επιτροπή

Βαρλάμης Ηρακλής

**Επίκουρος Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο**

Μιχαήλ Δημήτριος

**Επίκουρος Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο**

Αναγνωστόπουλος Δημοσθένης

**Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο**

Ο Παναγιώτης Κωνσταντίνος Γέμος,

δηλώνω υπεύθυνα ότι:

- 1) Είμαι ο κάτοχος των πνευματικών δικαιωμάτων της πρωτότυπης αυτής εργασίας και από όσο γνωρίζω η εργασία μου δε συκοφαντεί πρόσωπα, ούτε προσβάλει τα πνευματικά δικαιώματα τρίτων.
- 2) Αποδέχομαι ότι η ΒΚΠ μπορεί, χωρίς να αλλάξει το περιεχόμενο της εργασίας μου, να τη διαθέσει σε ηλεκτρονική μορφή μέσα από τη ψηφιακή Βιβλιοθήκη της, να την αντιγράψει σε οποιοδήποτε μέσο ή/και σε οποιοδήποτε μορφότυπο καθώς και να κρατά περισσότερα από ένα αντίγραφα για λόγους συντήρησης και ασφάλειας.

**Σε όλους όσους παραδίδουν την πτυχιακή
τους εργασία μετά την τελευταία στιγμή**

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τις ευχαριστίες μου ως προς τον επιβλέπων καθηγητή μου, κύριο Βαρλάμη Ηρακλή, για την ταχύτατη ανταπόκριση που είχα σε οποιοδήποτε ζήτημα σχετικά με την πτυχιακή μου εργασία αλλά και γενικά σε οτιδήποτε άλλο χρειαζόμουν τη γνώμη του και τη βοήθειά του. Θα ήθελα επίσης να ευχαριστήσω όλους τους καθηγητές και διδάσκοντες του Τμήματος Πληροφορικής και Τηλεματικής, οι οποίοι από το πρώτο έως και το τελευταίο έτος της φοίτησής μου, μου παρείχαν τα κατάλληλα ερεθίσματα και τις κατάλληλες γνώσεις, οι οποίες κρίθηκαν απαραίτητες για την εκπόνηση αυτής της εργασίας. Επιπλέον θα ήθελα να ευχαριστήσω τον Michael Nielsen (τον οποίο αν και δεν γνωρίζω προσωπικά) για το δωρεάν διαδικτυακό βιβλίο του “Neural Networks and Deep Learning”, Determination Press 2015, το οποίο με βοήθησε πάρα πολύ στο να καταλάβω τι είναι και πως λειτουργούν τα τεχνητά νευρωνικά δίκτυα.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περίληψη στα Ελληνικά	7
Abstract ή Περίληψη στα Αγγλικά	8
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	9
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	10
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ	11
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ	12
1. Εισαγωγή	13
2. Υπόβαθρο	16
2.1. Η Επανάσταση στην Κατηγοριοποίηση Εικόνας	16
2.2. Η Αρχιτεκτονική ενός Συνελικτικού Νευρωνικού Δικτύου	17
2.2.1 Επίπεδο Εισόδου	17
2.2.2 Συνελικτικό Επίπεδο	18
2.2.3 Συγκεντρωτικό Επίπεδο	20
2.2.4 Πλήρως Συνδεδεμένο Επίπεδο	21
2.2.5 Επίπεδο Εξόδου (Softmax)	21
2.3. Σύγχρονα Συνελικτικά Νευρωνικά Δίκτυα	23
2.3.1 Το Σύνολο Εικόνων ImageNet	23
2.3.2 Συνελικτικά Νευρωνικά Δίκτυα και ImageNet	24
2.3.3 Σημασία Αποτελεσμάτων στο ImageNet	26
3. Υλοποίηση	28
3.1. Περιγραφή των Δεδομένων	28
3.2: Χρησιμοποιώντας Κλασσική Μηχανική Μάθηση	30
3.2.1. Προεπεξεργασία Δεδομένων	30
3.2.2. Αλγόριθμοι	31
3.3: Χρησιμοποιώντας Ρηχά Νευρωνικά Δίκτυα	32
3.3.1. Αρχιτεκτονική Δικτύου BASE_0	32
3.3.2. Αρχιτεκτονική Δικτύου BASE_1	33
3.3.3. Προεπεξεργασία Δεδομένων	34
3.4: Χρησιμοποιώντας Deep Learning	36
3.4.1 Αρχιτεκτονική του Βαθέως Νευρωνικού Δικτύου CONV_1	36
3.4.2 Προεπεξεργασία Δεδομένων	38
3.4.3 Εκπαίδευση Μοντέλου	38

4. Αποτελέσματα	39
4.1. Κλασική μηχανική μάθηση	39
4.1.1 Αξιολόγηση στα Δεδομένα Εκπαίδευσης	39
4.1.2 Αξιολόγηση στα Δεδομένα Ελέγχου	41
4.2: Χρησιμοποιώντας Ρηχά Νευρωνικά Δίκτυα	42
4.2.1 Αποτελέσματα στα Δεδομένα Εκπαίδευσης	42
4.2.2 Αποτελέσματα στα Δεδομένα Ελέγχου	44
4.3. Χρησιμοποιώντας Deep Learning	44
4.3.1 Αποτελέσματα στα Δεδομένα Εκπαίδευσης	44
4.3.2 Αποτελέσματα στα Δεδομένα Ελέγχου	45
5. Συμπεράσματα - μελλοντικές επεκτάσεις	46
Παράρτημα Α: Αρχιτεκτονική Ρηχών Νευρωνικών Δικτύων	47
Παράρτημα Β: Λογισμικό Υλοποίησης Νευρωνικών Δικτύων	49
ΒΙΒΛΙΟΓΡΑΦΙΑ	51

Περίληψη στα Ελληνικά

Η ανάγκη για ανακάλυψη προτύπων σε δεδομένα καθώς και η κατηγοριοποίηση και η συσταδοποίηση διάφορων τύπων δεδομένων, ολοένα και αυξάνεται. Συνάμα με αυτή την ανάγκη, έρχεται και η ανάπτυξη της Τεχνητής Νοημοσύνης και ιδιαίτερα της Μηχανικής Μάθησης, όπου σκοπό έχει να ικανοποιήσει τις ανάγκες αυτές. Οι κλασσικοί αλγόριθμοι Μηχανικής Μάθησης έχουν αποδειχθεί αρκετά καλοί και αποδοτικοί στην εύρεση προτύπων σε δομημένα δεδομένα με γνωρίσματα. Πόσο καλά όμως αντιμετωπίζουν προβλήματα που αφορούν μη δομημένα δεδομένα; Δεδομένα στα οποία ο άνθρωπος έχει την έμφυτη ικανότητα να κατηγοριοποιεί και να ομαδοποιεί με μεγάλη αποδοτικότητα. Προβλήματα όπως η κατηγοριοποίηση μιας εικόνας, η αναγνώριση φωνής, η κατανόηση και η απόδοση ενός κειμένου, η δυνατότητα διαλόγου. Εκεί οι κλασσικές τεχνικές παρουσιάζουν χαμηλή απόδοση.

Αφού το μυαλό του ανθρώπου είναι καλό σε τέτοιες εργασίες, ίσως η λύση να κρύβεται στο να το προσομοιώσουμε. Τα Τεχνητά Νευρωνικά Δίκτυα, η λογική των οποίων βασίζεται στα νευρωνικά δίκτυα του εγκεφάλου, καθώς και οι νέες, πολυεπίπεδες αρχιτεκτονικές γνωστές και ως Deep Learning, απέδειξαν ότι όντως, προσομοιώνοντας τη λειτουργία των νευρωνικών κυττάρων, μπορούμε να επιτύχουμε υψηλές αποδόσεις σε αυτού του είδους τις εργασίες.

Σκοπός αυτής της πτυχιακής εργασίας είναι η ανάπτυξη τεχνητών νευρωνικών δικτύων για τη κατηγοριοποίηση εικόνων. Σε αυτή παρουσιάζονται διάφορες μέθοδοι και αρχιτεκτονικές τεχνητών νευρωνικών δικτύων, ξεκινώντας από απλές και καταλήγοντας σε πιο σύνθετες, εστιάζοντας στο τρόπο λειτουργίας τους, τις προκλήσεις τους καθώς και στην αποδοτικότητά τους. Επίσης, αποδεικνύεται η ανωτερότητα των τεχνικών Deep Learning έναντι των ρηχών αρχιτεκτονικών τεχνητών νευρωνικών δικτύων, καθώς και γενικά η ανωτερότητα των τεχνητών νευρωνικών δικτύων έναντι των κλασσικών μεθόδων Μηχανικής Μάθησης.

Λέξεις κλειδιά: Κατηγοριοποίηση Εικόνων, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Τεχνητά Νευρωνικά Δίκτυα, Deep Learning.

Abstract ή Περίληψη στα Αγγλικά

The need for discovering patterns in data and the classification and clustering of various data types, is growing faster and faster. Along with this need, comes Artificial Intelligence and especially Machine Learning the purpose of which is to satisfy these needs. The classical Machine Learning algorithms have proven to be extremely good and efficient in finding patterns in featured data. How well do they confront problems regarding unstructured data? Data in which humans have the born ability to classify and clusterify with great performance. Problems like image classification, speech recognition, text understanding, text composing and chatting. In these situations, the classical techniques presents low performance.

Since the human brain is good at these tasks, maybe the solution is to simulate it. Artificial Neural Networks, which are based on the neural networks of the brain, and the new multi-layer architectures known as Deep Learning, have proven that indeed, by simulating the functionality of the neuron cells, we can achieve high performance in these kind of tasks.

The purpose of this thesis, is the development of artificial neural networks for image classification. It presents different methods and architectures of neural networks, beginning with the simplest and ending with the most complex, focusing in their way of functioning, their challenges and their efficiency. In addition, the superiority of Deep Learning techniques is proven over the shallow neural networks, as well as the superiority that the neural networks, even the most simple ones, present over the classical Machine Learning techniques, in image classification.

Keywords: Image Classification, Artificial Intelligence, Machine Learning, Artificial Neural Networks, Deep Learning.

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικ. 2.3.1.1 Κομμάτι από τις εικόνες του ImageNet	Σελ. 23
Εικ. 2.3.2.1 Αρχιτεκτονική του CNN των Krizhevsky et. al	Σελ. 24
Εικ. 2.3.3.1 Αντιφατικές Εικόνες	Σελ. 27
Εικ. A.1 Νευρώνας	Σελ. 47
Εικ. A.2 Δίκτυο από νευρώνες	Σελ. 47

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πιν. 3.3.3.1 Παράδειγμα onehot διανύσματος

Σελ. 35

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχ. 2.2.1.1 Νευρώνες Εισόδου	Σελ. 17
Σχ. 2.2.2.1 Εφαρμογή Φίλτρου Συνέλιξης	Σελ. 18
Σχ. 2.2.2.2 Δημιουργία Πρώτου Κρυφού Επιπέδου	Σελ. 19
Σχ. 2.2.2.3 Feature Maps	Σελ. 20
Σχ. 2.2.3.1 Επίπεδα Εισόδου, Συνέλιξης και Συγκέντρωσης	Σελ. 20
Σχ. 2.2.5.1 Ολοκληρωμένο Συνελικτικό Νευρωνικό Δίκτυο	Σελ. 22
Σχ. 2.3.2.2 Dropout	Σελ. 25
Σχ. 3.1.1.1 Κατανομή των εικόνων ανά κατηγορία	Σελ. 29
Σχ. 3.3.1.2 Αρχιτεκτονική Μοντέλου BASE_0	Σελ. 33
Σχ. 3.3.2.3 Αρχιτεκτονική Μοντέλου BASE_1	Σελ. 33
Σχ. 3.4.1.1 Αρχιτεκτονική Μοντέλου CONV_1	Σελ. 37
Σχ. 4.2.1.1 Μοντέλο BASE_0	Σελ. 42
Σχ. 4.2.1.2 Μοντέλο BASE_1	Σελ. 43
Σχ. 4.3.1.1 Μοντέλο CONV_1	Σελ. 45

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

CSV	Comma Separated Values
GPU	Graphics Processing Unit
TNN	Τεχνητά Νευρωνικά Δίκτυα
CNN	Convolutional Neural Network

1. Εισαγωγή

Τα τεχνητά νευρωνικά δίκτυα (TNN) αποτελούν αντικείμενο ενδιαφέροντος της Τεχνητής Νοημοσύνης και συγκεκριμένα της Μηχανικής Μάθησης. Υπάρχουν ως ιδέα από τα μέσα του προηγούμενου αιώνα [1]. Η έρευνα πάνω σε αυτά από τότε έως και τα τέλη του 20 αιώνα, συνέχισε με σταθερούς ρυθμούς. Παρ' όλα αυτά, κατά την περίοδο εκείνη εφευρέθηκαν μέθοδοι και τεχνικές που χρησιμοποιούνται κατά πολύ στις σύγχρονες αρχιτεκτονικές νευρωνικών δικτύων. Όμως δεν υπήρξαν πρακτικές εφαρμογές που να τις χρησιμοποιούν. Παράδειγμα τέτοιων σημαντικών μεθόδων, αποτελούν η εφεύρεση της μεθόδου Back-propagation [2] και της μεθόδου Stochastic Gradient Descent [3]. Μάλιστα, οι δύο αυτές μέθοδοι μαζί, αποτελούν μέχρι και σήμερα τον κύριο πυρήνα των περισσότερων αρχιτεκτονικών νευρωνικών δικτύων.

Η σταθερότητα στην έρευνα των νευρωνικών δικτύων, έσπασε στις αρχές του 2000, όπου και εκτινάχθηκε στα ύψη, με το αποκορύφωμα να σημειώνεται την τρέχουσα δεκαετία και αναμένεται να συνεχίζει σταθερά με αυτό τον υψηλό ρυθμό στο άμεσο μέλλον. Η έρευνα αυτή έθεσε το κατάλληλο έδαφος προκειμένου τα TNN να χρησιμοποιηθούν σε πρακτικές εφαρμογές. Η αναγνώριση φωνής, τα διαδικτυακά ρομπότ-διάλογοι (chatbots) και τα αυτο-οδηγούμενα αυτοκίνητα, οφείλουν σε τεράστιο ποσοστό την επιτυχία τους στα TNN. Η ραγδαία αυτή ανάπτυξη, οφείλεται σε μεγάλο βαθμό στην τρομερή αύξηση της υπολογιστικής ισχύος που χαρακτήρισε την περίοδο αυτή, επιτρέποντας πλέον την διεξαγωγή των άλλοτε χρονοβόρων και κοστοβόρων πειραμάτων σε πολύ λίγο χρόνο και με πολύ λίγο κόστος. Το ίδιο ισχύει και για τον χρόνο εκπαίδευσης των δικτύων, ο οποίος μειώθηκε, σε επίπεδα όπου κάθε σύγχρονος υπολογιστής μπορεί να εκπαιδεύσει ένα νευρωνικό δίκτυο σε λογικά χρονικά περιθώρια.

Οι σύγχρονες τεχνικές αυτές, που είναι ευρύτερα γνωστές πλέον ως τεχνικές Deep Learning, κατάφεραν να πετύχουν εξαιρετικές επιδόσεις σε πολλούς τομείς, σε αντίθεση με τις παλιές, ρηχές αρχιτεκτονικές νευρωνικών δικτύων (από εδώ και στο εξής οι όροι Τεχνητά Νευρωνικά Δίκτυα και Νευρωνικά Δίκτυα θα έχουν τη ίδια

σημασία). Πλέον υπάρχουν δεκάδες πολυεπίπεδες αρχιτεκτονικές, όπου η κάθε μία είναι καλή σε ένα συγκεκριμένο τομέα. Μία είναι όμως αυτή η οποία κέντρισε το μεγαλύτερο ενδιαφέρον εξ αυτών. Αυτή είναι η αρχιτεκτονική των Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks - CNN), τα οποία κατάφεραν να αποδώσουν εξαιρετικά σε μία από τις πιο δύσκολες προκλήσεις της Τεχνητής Νοημοσύνης, την αναγνώριση εικόνων.

Το πρώτο CNN το οποίο αποτέλεσε τεράστια επιρροή για τα μοντέρνα συνελικτικά νευρωνικά δίκτυα ήταν το επιστημονικό έγγραφο των LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. το 1998, ονόματι: “Gradient-based learning applied to document recognition” [4], αν και η ιδέα των CNN υπήρχε από την δεκαετία του 1970. Σε αυτό το έγγραφο, αναφέρεται μία αρχιτεκτονική συνελικτικού νευρωνικού δικτύου όπου σκοπό έχει την αναγνώριση χειρόγραφων ψηφίων. Το συγκεκριμένο μάλιστα δίκτυο νευρώνων, είχε εφαρμογή στον εμπορικό τομέα, όπου χρησιμοποιήθηκε στην αυτοματοποιημένη ανάγνωση εκατομμυρίων επιταγών.

Η ανάγνωση χαρακτήρων είναι μία πρόκληση. Δεν είναι όμως από τις πιο ισχυρές στον τομέα της. Μία από τις μεγαλύτερες προκλήσεις που καλούνται να επιλύσουν τα CNN είναι η αναγνώριση πολλαπλών αντικειμένων σε μία εικόνα, από ένα τεράστιο εύρος πιθανών αντικειμένων. Μια γνωστή τέτοια πρόκληση αποτελεί η κατηγοριοποίηση του συνόλου εικόνων από το ImageNet. Το ImageNet αποτελεί μία τεράστια βιβλιοθήκη εικόνων οι οποίες είναι κατηγοριοποιημένες από τον άνθρωπο. Μέχρι στιγμής απαριθμεί 14 εκατομμύρια φωτογραφίες σε 20 χιλιάδες κατηγορίες. Από τον όγκο των δεδομένων και των πιθανών κατηγοριών και μόνο, μπορεί κάποιος να αντιληφθεί τη δυσκολία του προβλήματος αυτού.

Το 1990, όπου οι υπολογιστές δεν είχαν την υψηλή υπολογιστική ισχύ που έχουν σήμερα, χρησιμοποιούνταν οι κλασσικές τεχνικές Μηχανικής Μάθησης, σε προβλήματα αναγνώρισης εικόνας. Οι τεχνικές αυτές απαιτούσαν για το σύνολο των εικόνων, να διέλθουν από αρκετές φάσεις προεπεξεργασίας, προκειμένου να μπορέσουν να λειτουργήσουν με σχετικά καλή απόδοση. Τα νευρωνικά δίκτυα σε αντίθεση, δεν χρειάζονται ιδιαίτερη προεπεξεργασία στα δεδομένα εισόδου. Επίσης,

συνήθως εφαρμόζονταν πάνω σε γνωρίσματα τα οποία εξάγονταν από τις εικόνες κατά τη διάρκεια της προεπεξεργασίας. Υπερνικούν τα νευρωνικά δίκτυα μόνο όσον αφορά τον χρόνο εκπαίδευσης και την απλότητά τους.

Σε αυτή την πτυχιακή εργασία, παρουσιάζονται μέθοδοι κατηγοριοποίησης εικόνων, χρησιμοποιώντας πολυεπίπεδα και ρηχά νευρωνικά δίκτυα. Επίσης, συγκρίνεται η απόδοση τους σε σχέση με κλασσικές τεχνικές Μηχανικής Μάθησης. Ως σύνολο δεδομένων, χρησιμοποιούνται οι εικόνες από τον διαγωνισμό της διαδικτυακής ιστοσελίδας Kaggle: “The Nature Conservancy Fisheries Monitoring” [5], όπου αποτελούνται από στιγμιότυπα που έχουν ληφθεί από κάμερες πάνω σε αλιευτικά πλοία. Σκοπός του διαγωνισμού είναι η κατηγοριοποίηση των εικόνων αυτών, βάσει του τύπου των ψαριών που μπορεί να βρίσκονται στην εικόνα. Υπάρχουν 8 συνολικά κατηγορίες, οι 7 από τις οποίες αφορούν είδη ψαριών και μία κατηγορία η οποία περιλαμβάνει το ενδεχόμενο να μην υπάρχει ψάρι στην εικόνα.

2. Υπόβαθρο

Οι προοπτικές που υπόσχονται τα νευρωνικά δίκτυα σε εργασίες τις οποίες ο άνθρωπος είναι καλός στο να εκτελεί, ώθησαν την έρευνα σε πολύ σημαντικές ανακαλύψεις. Εστιάζοντας σε μία από αυτές τις εργασίες, στην ικανότητα αναγνώρισης εικόνων του ανθρώπου, τα τεχνητά νευρωνικά δίκτυα προσπαθούν να πετύχουν αποτελέσματα κοντά στην ανθρώπινη επίδοση. Σύγχρονες έρευνες, κατάφεραν να δημιουργήσουν νευρωνικά δίκτυα κατηγοριοποίησης εικόνων, που όχι μόνο πλησιάζουν τον άνθρωπο στις επιδόσεις, αλλά και τον ξεπερνούν.

2.1. Η Επανάσταση στην Κατηγοριοποίηση Εικόνας

Το πρώτο νευρωνικό δίκτυο το οποίο έφερε μεγάλη επανάσταση στην κατηγοριοποίηση εικόνων ήταν το δίκτυο το οποίο αναπτύχθηκε από τους Yann LeCun et. al, το οποίο αναφέρεται στο επιστημονικό έγγραφο “Gradient-based learning applied to document recognition” [4]. Σε αυτό το έγγραφο εδραιώθηκαν οι βάσεις, για την δημιουργία νευρωνικών δικτύων τα οποία μιμούνται τη διασύνδεση των βιολογικών νευρωνικών δικτύων που χρησιμοποιούνται για την όραση. Βάσει αυτής της έρευνας καθιερώθηκαν ιδέες που χρησιμοποιούνται ακόμη και σήμερα στα σύγχρονα συνελκτικά νευρωνικά δίκτυα. Η πιο σημαντική εκ αυτών ήταν η χρήση τριών ειδών επιπέδων νευρωνικών δικτύων: το επίπεδο συνέλιξης (convolutional layers), το επίπεδο συγκέντρωσης (pooling layer) και το πλήρως συνδεδεμένο επίπεδο (fully connected layer). Ο συνδυασμός αυτών των τύπων επιπέδων, χρησιμοποιείται έως σήμερα για την κατασκευή σύγχρονων συνελκτικών δικτύων.

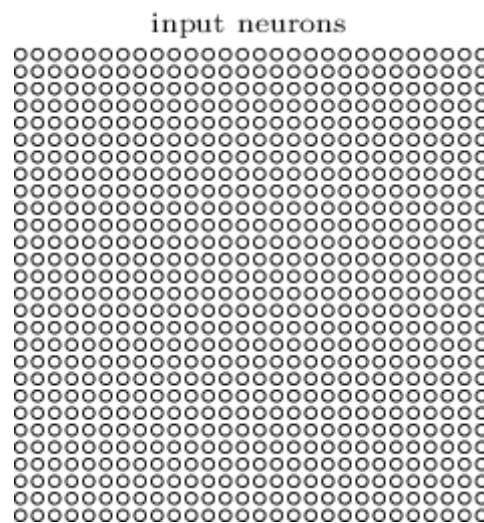
Η κατασκευή πλέον ενός συνελκτικού νευρωνικού δικτύου, περιλαμβάνει τον ορισμό του πλήθους, του τύπου και του μεγέθους των επιπέδων, καθώς επίσης της μεταξύ τους διάταξης, του ορισμού της συνάρτησης συγκέντρωσης και του μεγέθους φίλτρου αυτής και του ορισμού μεγέθους του φίλτρου συνέλιξης. Οι κατάλληλες τιμές των παραπάνω υπερ-παραμέτρων, θα πρέπει να βρεθούν μέσω μίας διαδικασίας

δοκιμής και λάθους (trial and error). Δεν υπάρχει ακόμα κάποιος αυστηρός και πλήρως ορισμένος τρόπος για την εύρεση αυτών των παραμέτρων, παρά μόνο ευριστικές μέθοδοι, οι οποίες βασίζονται στην εμπειρία.

2.2. Η Αρχιτεκτονική ενός Συνελικτικού Νευρωνικού Δικτύου

2.2.1 Επίπεδο Εισόδου

Τα συνελκτικά νευρωνικά δίκτυα δέχονται ως είσοδο στο πρώτο επίπεδο νευρώνων, εικόνες οποιαδήποτε διάστασης. Συνήθως χρησιμοποιούνται για εικόνες δύο διαστάσεων, αλλά μπορούν να δοθούν και δεδομένα τριών διαστάσεων [6]. Αυτό είναι ένα θετικό των δικτύων αυτών έναντι των ρηχών αρχιτεκτονικών, οι οποίες συνήθως δεν υπολογίζουν τις διαστάσεις της εικόνας, χάνοντας την σημαντική πληροφορία την οποία προσφέρουν. Επομένως, συνήθως ως είσοδο έχουμε έναν πίνακα δύο διαστάσεων από όπου κάθε τιμή του πίνακα αντιστοιχεί σε ένα εικονοστοιχείο.

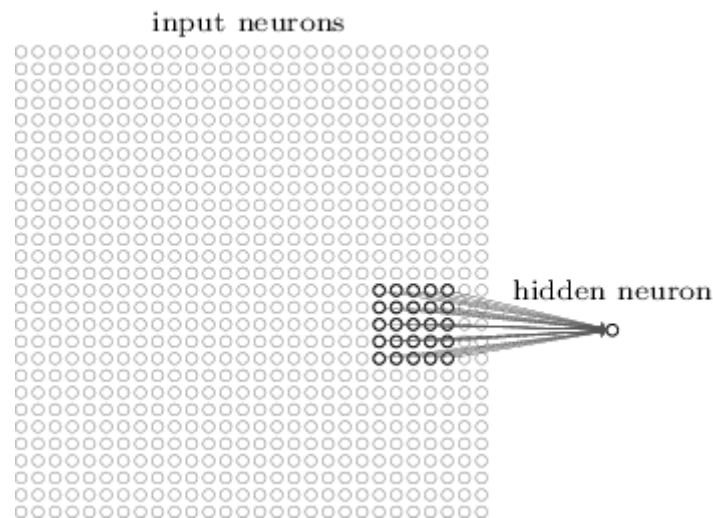


Νευρώνες Εισόδου (Σχ. 2.2.1.1)

Στο Σχ. 2.2.1.1, φαίνεται η διάταξη ενός επιπέδου εισόδου μεγέθους $N \times N$. Σε περίπτωση που έχουμε χρωματική εικόνα, τότε η είσοδος αποτελείται από τρία τέτοια κανάλια.

2.2.2 Συνελικτικό Επίπεδο

Το επόμενο επίπεδο είναι συνήθως συνελικτικό. Το συνελικτικό επίπεδο, χρησιμοποιεί φίλτρο μεγέθους $M \times M$ (γνωστό και ως local receptive field) το οποίο εφαρμόζεται, μετακινούμενο ανά s το πλήθος pixel (ονόματι strides) πάνω στην εικόνα.



Εφαρμογή Φίλτρου Συνέλιξης (Σχ. 2.2.2.1)

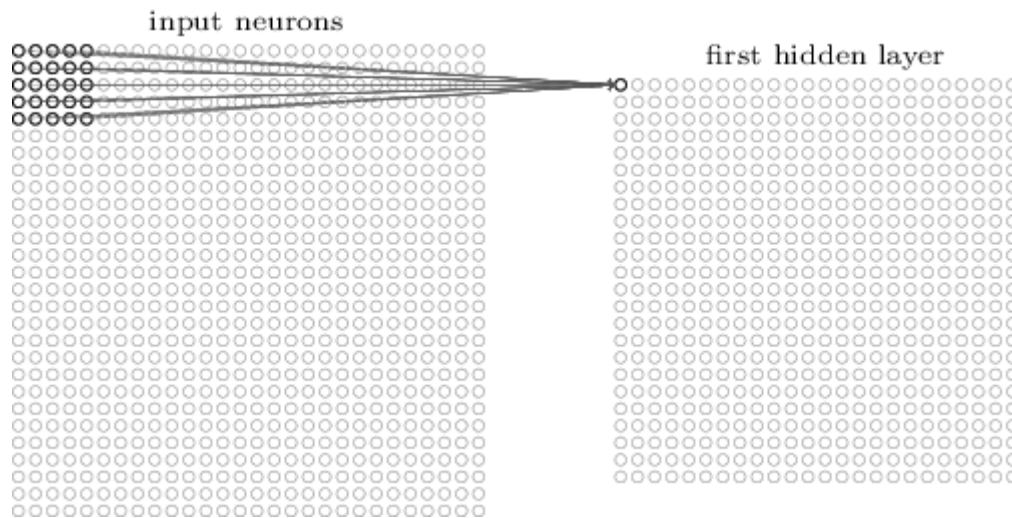
Για την τιμή του κάθε νευρώνα στο πρώτο κρυφό επίπεδο, υπολογίζεται το άθροισμα των τιμών του φίλτρου, τα οποία ονομάζονται βάρη, προστιθέμενα από μία σταθερή ποσότητα, την κλίση (bias). Ύστερα αυτό το άθροισμα, οδηγείται σε μία συνάρτηση ενεργοποίησης του νευρώνα, η έξοδος της οποίας αποτελεί την τιμή του νευρώνα. Υπάρχουν πολλές συναρτήσεις ενεργοποίησης, αλλά η πιο σύνηθες για τα συνελικτικά δίκτυα, είναι η συνάρτηση ανόρθωσης. Στη περίπτωση αυτή μάλιστα, οι νευρώνες ονομάζονται ανορθωτικές γραμμικές μονάδες (linear rectified units - ReLU). Για τον υπολογισμό της τιμής του κάθε νευρώνα, χρησιμοποιούνται τα ίδια βάρη και κλίση. Για το λόγο αυτό, ορίζονται ως κοινά βάρη και κλίση (shared weights and bias).

Το αποτέλεσμα του επιπέδου συνέλιξης είναι να δημιουργηθεί το πρώτο κρυφό επίπεδο το οποίο αποτελείται από $K \times K$ νευρώνες, όπου:

$$K = (N - M) / s + 1$$

Η πιο γενική φόρμουλα, περιλαμβάνει το P , όπου είναι το μέγεθος γεμίσματος μηδενικών (zero padding) στις άκρες του πίνακα:

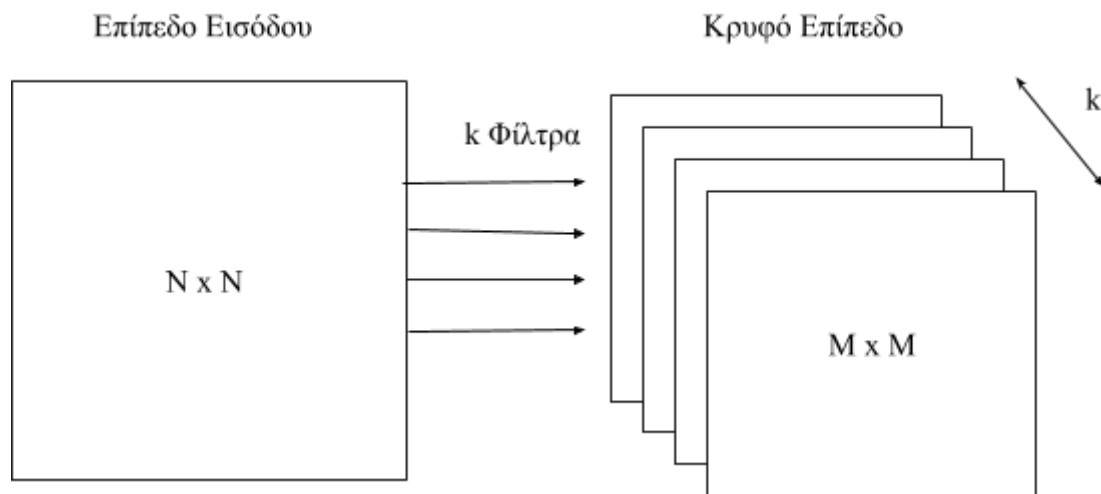
$$K = (N - M + 2P) / s + 1$$



Δημιουργία Πρώτου Κρυφού Επιπέδου (Σχ. 2.2.2.2)

Επειδή ο κάθε νευρώνας του κρυφού επιπέδου, υπολογίζεται μέσω ενός φίλτρου με κοινά βάρη και κλίση, για κάθε ζευγάρι εικόνας εισόδου και φίλτρου, υπολογίζεται ένα κρυφό επίπεδο, το οποίο παρουσιάζει ένα συγκεκριμένο χαρακτηριστικό. Η αντιστοίχιση μεταξύ του συνδυασμού εισόδου και φίλτρου προς το κρυφό επίπεδο το οποίο αναπαριστά το χαρακτηριστικό, ονομάζεται feature-map.

Σε κάθε συνελκτικό επίπεδο μπορούμε να έχουμε παραπάνω από ένα φίλτρα, με αποτέλεσμα να παίρνουμε ως έξοδο τόσα χαρακτηριστικά όσα και τα φίλτρα, αυξάνοντας έτσι τα feature maps. Τα χαρακτηριστικά αυτά, θεωρούνται ως ένα κρυφό επίπεδο το οποίο περιέχει πολλαπλά κανάλια, όπου ένα κανάλι ισούται με ένα χαρακτηριστικό.



Feature Maps (Σχ. 2.2.2.3)

2.2.3 Συγκεντρωτικό Επίπεδο

Έπειτα από τα συνελκτικά επίπεδα, συνήθως ακολουθούν τα συγκεντρωτικά επίπεδα, σκοπός των οποίων είναι η απλοποίηση της πληροφορίας του συνελκτικού επιπέδου.



Επίπεδα Εισόδου, Συνέλιξης και Συγκέντρωσης (Σχ. 2.2.3.1)

Τα συγκεντρωτικά επίπεδα, λειτουργούν με όμοια λογική με τα συνελικτικά επίπεδα, δηλαδή με φίλτρα μεγέθους $M \times M$ μονάχα που εφαρμόζονται ανά $s = M$ στοιχεία. Επομένως, για είσοδο μεγέθους $N \times N$, εξάγουν έξοδο μεγέθους $K \times K$, όπου $K = N/M$. Επίσης, τα φίλτρα, αντί για σύνολο τιμών από βάρη, αποτελούν συναρτήσεις οι οποίες εφαρμόζονται ανά το σύνολο στοιχείων της κάθε περιοχής που εφαρμόζεται το φίλτρο. Η πιο σύνηθες συνάρτηση για αυτό το σκοπό, είναι η συνάρτηση μεγίστου. Χρησιμοποιώντας αυτή, η συγκέντρωση είναι γνωστή στη βιβλιογραφία με το όνομα max-pooling.

2.2.4 Πλήρως Συνδεδεμένο Επίπεδο

Στα τελευταία κρυφό επίπεδο των συνελικτικών νευρωνικών δικτύων, κατά κανόνα τοποθετείται πλήρης συνδεδεμένο επίπεδο νευρώνων. Αυτά λειτουργούν όπως τα ρηγά νευρωνικά δίκτυα (βλ. Παράρτημα Α), όπου κάθε νευρώνας του ενός επιπέδου είναι συνδεδεμένος με κάθε νευρώνα του επόμενου επιπέδου. Συνήθως αποτελούνται από ReLU νευρώνες, αλλά χρησιμοποιούνται και άλλες συναρτήσεις ενεργοποίησης. Είναι υπεύθυνα για την σύνδεση των νευρώνων εξόδου με την έξοδο των συγκεντρωτικών νευρώνων του προηγούμενου επιπέδου. Διαισθητικά, τα συνελικτικά και τα συγκεντρωτικά επίπεδα, συγκεντρώνουν την χωρική πληροφορία που κρύβεται στην εικόνα, ενώ τα πλήρη συνδεδεμένα επίπεδα, χρησιμοποιούν την αφηρημένη αυτή πληροφορία, πιο σφαιρικά, προκειμένου να συγκεντρώσουν στοιχεία για τη σωστή κατηγορία ταξινόμησης.

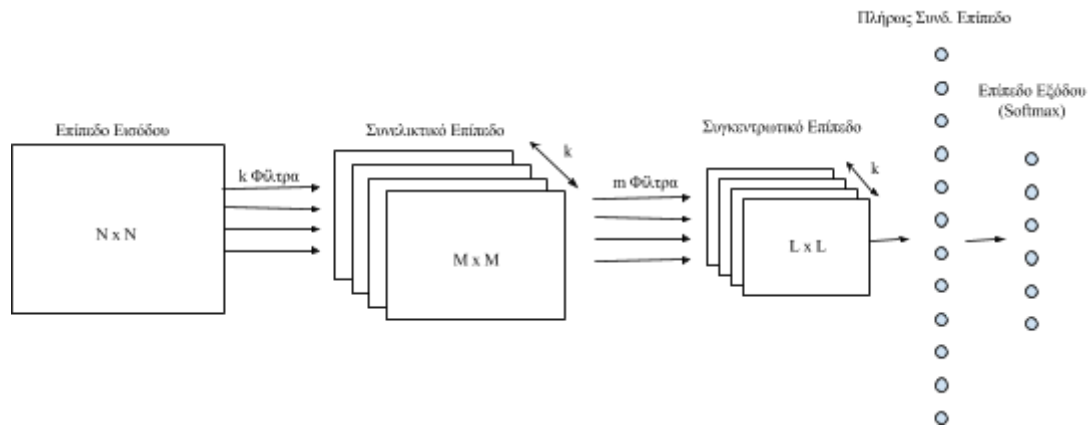
2.2.5 Επίπεδο Εξόδου (Softmax)

Το τελευταίο επίπεδο εξόδου, συνήθως είναι το επίπεδο softmax. Το επίπεδο αυτό έχει τόσους νευρώνες, όσες είναι και οι διάφορες κατηγορίες των εικόνων. Είναι υπεύθυνο στο να μετατρέπει τις τιμές που φτάνουν σε κάθε νευρώνα εξόδου σε πιθανότητες, προκειμένου ύστερα να χρησιμοποιηθούν για την αξιολόγηση του μοντέλου ή την πρόβλεψη της κατηγορίας μιας άγνωστης εικόνας πιθανοτικά. Η πρόβλεψη μπορεί να γίνει και ντετερμινιστικά, προβλέποντας ως τη σωστή κατηγορία, αυτή η οποία έχει τη μεγαλύτερη πιθανότητα. Το επίπεδο softmax λειτουργεί εφαρμόζοντας την συνάρτηση softmax ή αλλιώς τυποποιημένη εκθετική συνάρτηση στις τιμές των νευρώνων εξόδου. Η συνάρτηση αυτή έχει τύπο:

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

Όπου, z_j είναι η τιμή του j νευρώνα εξόδου.

Στην παρακάτω εικόνα φαίνεται ολοκληρωμένο το συνελκτικό δίκτυο, όπου παρουσιάζονται όλοι οι τύποι επιπέδων νευρώνων που χρησιμοποιούνται σε τέτοια είδους δίκτυα. Στην εικόνα αυτή παρουσιάζεται η πιο απλή αρχιτεκτονική ενός τέτοιου δικτύου, καθώς οι σύγχρονες αρχιτεκτονικές χρησιμοποιούν περισσότερα από ένα κρυφά επίπεδα, αποσκοπώντας μεγαλύτερη αποδοτικότητα, έχοντας πολλαπλά συνελκτικά, συγκεντρωτικά και πλήρως συνδεδεμένα επίπεδα.



Ολοκληρωμένο Συνελκτικό Νευρωνικό Δίκτυο (Σχ. 2.2.5.1)

2.3. Σύγχρονα Συνελικτικά Νευρωνικά Δίκτυα

Βασιζόμενοι στα τρία βασικά συστατικά (συνελικτικό, συγκεντρωτικό και πλήρως συνδεδεμένο επίπεδο) των συνελικτικών νευρωνικών δικτύων, κατασκευάστηκαν αρχιτεκτονικές την τελευταία δεκαετία οι οποίες επέκτειναν την εφαρμογή των δικτύων αυτών από το απλό σύνολο δεδομένων χειρόγραφων γραμμάτων, σε πιο αφηρημένες και πολύπλοκες έννοιες. Πλέον οι προκλήσεις που αντιμετωπίζουν τα συνελικτικά νευρωνικά δίκτυα, αφορούν την κατηγοριοποίηση εικόνων, από ένα τεράστιο σύνολο εικόνων, χιλιάδων ετικετών, οι οποίες αφορούν διάφορα είδη περιεχομένου.

2.3.1 Το Σύνολο Εικόνων ImageNet

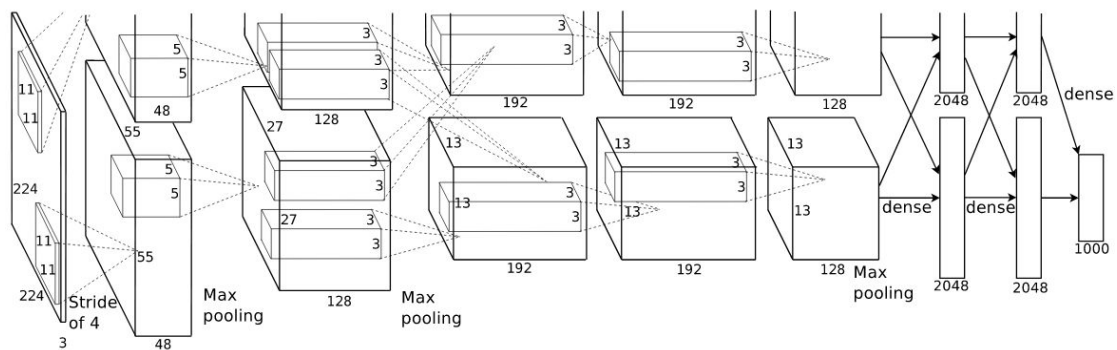
Ένα από τα πιο γνωστά σύνολα δεδομένων τα οποία περιέχουν εικόνες από χιλιάδες κατηγορίες, είναι το ImageNet [7]. Οι εικόνες στο ImageNet χρησιμοποιούνται από πολλούς ερευνητές ως σημείο αναφοράς (benchmark) για τις αρχιτεκτονικές νευρωνικών δικτύων που αναπτύσσουν. Μέχρι στιγμής αποτελείται από περίπου 14,2 εκατομμύρια χρωματικές εικόνες, οι οποίες έχουν συλλεγεί από το ανοικτό διαδίκτυο, σε 22 χιλιάδες κατηγορίες και τα νούμερα αυξάνονται. Πρόκειται για ένα σύνολο δεδομένων το οποίο χρειάστηκε πολλούς ανθρώπους προκειμένου να κατηγοριοποιήσουν σωστά τις εικόνες, ώστε να αντιστοιχιστεί η σωστή ετικέτα για την κάθε εικόνα. Για τον σκοπό αυτό χρησιμοποιήθηκε η crowdsourcing πλατφόρμα Amazon Mechanical Turk



Κομμάτι από τις εικόνες του ImageNet (Εικ. 2.3.1.1)

2.3.2 Συνελκτικά Νευρωνικά Δίκτυα και ImageNet

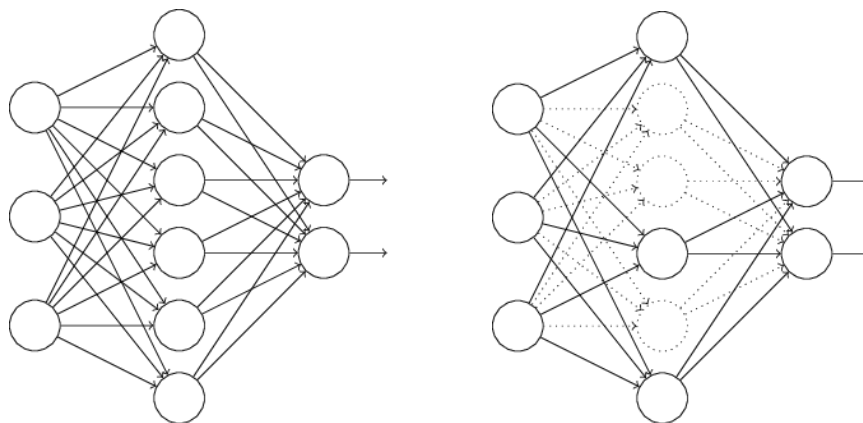
Υπήρξαν πολλές προσπάθειες από διάφορους ερευνητές στην κατασκευή συνελκτικών νευρωνικών δικτύων με σκοπό την κατηγοριοποίηση των εικόνων από το ImageNet. Από τις πρώτες, αξιόλογες επιστημονικές εργασίες που κατάφεραν να πετύχουν ακρίβεια αρκετά υψηλή για την εποχή τους, ήταν αυτή των Krizhevsky et. al [8] που χρησιμοποίησε συνελκτικά νευρωνικά δίκτυα στην κατηγοριοποίηση εικόνων του ImageNet. Το νευρωνικό δίκτυο που αναπτύχθηκε από τους ερευνητές αυτού του επιστημονικού εγγράφου, ήταν ένα δίκτυο 60 εκατομμυρίων παραμέτρων και 650 χιλιάδων νευρώνων. Αποτελούνταν από 5 συνελκτικά επίπεδα ακολουθούμενα από συγκεντρωτικά επίπεδα που χρησιμοποιούσαν την συνάρτηση μεγίστου, με τα τελευταία επίπεδα να αποτελούνται από τρία πλήρως διασυνδεδεμένα επίπεδα και ένα επίπεδο εξόδου softmax. Αποτελούσε από τις πρώτες αρχιτεκτονικές οι οποίες χρησιμοποιούσαν την επεξεργαστική ισχύ και μνήμη της κάρτας γραφικών (GPU), με σκοπό την ταχύτερη εκπαίδευση των μοντέλων νευρωνικών δικτύων.



Αρχιτεκτονική του συνελκτικού νευρωνικού δικτύου των Krizhevsky et. al (Εικ.

2.3.2.1)

Για τη μείωση της υπερ-εκπαίδευσης (overfitting), χρησιμοποιήθηκε μία νέα τεχνική γνωστή ως dropout. Το dropout, είναι μία μοντέρνα τεχνική κανονικοποίησης των νευρωνικών δικτύων, η οποία λειτουργεί θέτοντας μερικούς νευρώνες εκτός λειτουργίας κατά την διάρκεια της εκπαίδευσης. Με αυτόν τον τρόπο, σε κάθε μικρό σετ εκπαίδευσης (mini-batch), λαμβάνονται διαφορετικά νευρωνικά δίκτυα, το αποτέλεσμα των οποίων συμψηφίζεται, προκειμένου το τελικό αποτέλεσμα να είναι η κατηγορία με τις περισσότερες ψήφους.



Dropout. Αριστερή και δεξιά εικόνα αποτελούν τμήματα νευρωνικού δικτύου πριν και μετά την εφαρμογή της μεθόδου dropout αντίστοιχα (Σχ. 2.3.2.2)

Κατά αυτόν τον τρόπο είναι γνωστό από τις κλασσικές μεθόδους Μηχανικής Μάθησης, ότι αποφεύγεται η υπερ-εκπαίδευση, καθώς συνήθως η μειοψηφία αποτελείται από τα overfitted μοντέλα (στη περίπτωση μας, μοντέλα νευρωνικών δικτύων). Στο επιστημονικό έγγραφο, αναφέρεται ότι η μέθοδος αυτή ήταν εξαιρετικά αποδοτική.

Παραλλαγή του μοντέλου αυτού, χρησιμοποιήθηκε στον διαγωνισμό ILSVRC-2012, όπου και κέρδισε την πρώτη θέση, πετυχαίνοντας top-5 ρυθμό σφαλμάτων (top-5 error rate) 15.3%. Ο ρυθμός αυτός θεωρεί λανθασμένη την κατηγοριοποίηση της εικόνας, αν δεν βρίσκεται εντός των 5 πιο πιθανών προβλέψεων. Το αποτέλεσμα αυτό ήταν πολύ μεγαλύτερο από αυτό της ομάδας που ήταν στην δεύτερη θέση, όπου πέτυχε 26.2% top-5 ρυθμό σφαλμάτων.

Πιο σύγχρονες αρχιτεκτονικές, όπως η GoogLeNet [9], κατάφερε να πετύχει top-5 ρυθμό σφάλματος 6.67%, αποδίδοντας παραπάνω από το διπλάσιο από ότι το συνελκτικό νευρωνικό δίκτυο των Krizhevsky et. al. Το όνομά της το πήρε προς τιμήν της πρωτοποριακής αρχιτεκτονικής LeNet-5 των Yann LeCun et. al. [11]. Η αρχιτεκτονική αυτή χρησιμοποιούσε 12 επίπεδα νευρώνων, και κατάφερε να πετύχει αυτό το χαμηλό ρυθμό σφάλματος, λόγω μίας βελτιστοποίησης της αρχιτεκτονικής, όπου επέτρεπε την κλιμάκωση του βάθους και του πλάτους του δικτύου, διατηρώντας σταθερή την υπολογιστική πολυπλοκότητα.

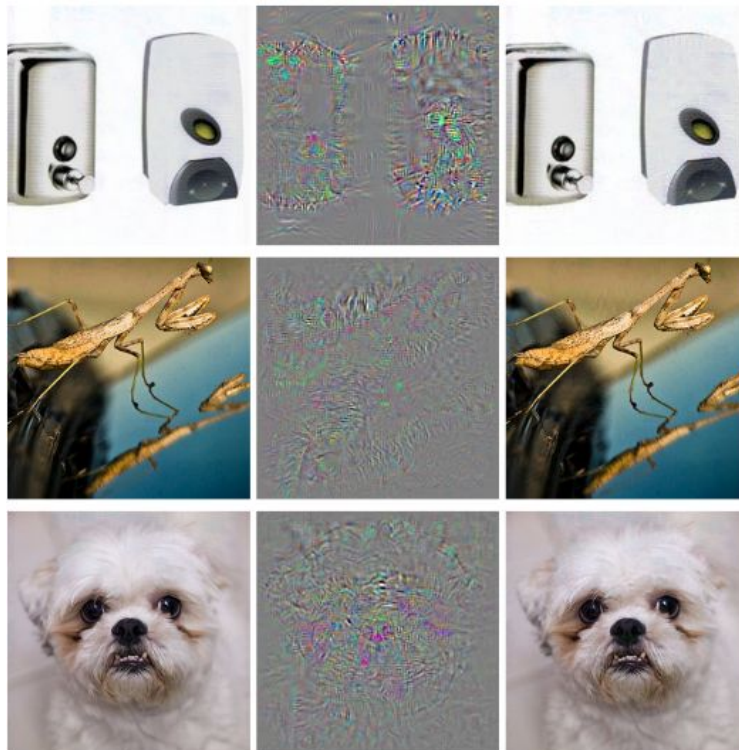
2.3.3 Σημασία Αποτελεσμάτων στο ImageNet

Ο νικητής του πιο πρόσφατου διαγωνισμού, ήταν η ομάδα Trimps-Soushen, όπου πέτυχε top-5 ρυθμό σφάλματος 2,991%. Ποσοστό το οποίο ξεπερνάει ακόμα και τον άνθρωπο [12]. Τι σημασία έχει όμως αυτό το ποσοστό; Έχει ξεπεράσει όντως η μηχανή τον άνθρωπο; Η αλήθεια είναι ότι το πρόβλημα που εισάγει η κατηγοριοποίηση του συνόλου εικόνων ImageNet είναι αρκετά περιορισμένο και δεν είναι αντιπροσωπευτικό των εικόνων που υπάρχουν σε άλλες εφαρμογές. Ούτε επίσης ο top-5 ρυθμός σφάλματος είναι πλήρως κατατοπιστικός.

Πέρα από το ImageNet, υπάρχουν και άλλα γνωστά σύνολα εικόνων, όπως το σύνολο εικόνων από το Google Street View. Η πρόκληση σε αυτό το σύνολο εικόνων, είναι η αναγνώριση των αριθμών οδών από τις εικόνες. Μία ομάδα από τη Google, χρησιμοποίησε ένα συνελκτικό νευρωνικό δίκτυο, το οποίο κατάφερε να αναγνωρίσει 100 εκατομμύρια αριθμούς οδών, με ακρίβεια ίδια με αυτή ενός ανθρώπου [13]. Στο επιστημονικό τους έγγραφο αναφέρουν ότι ή μέθοδος αυτή, πιστεύουν, ότι έλυσε το πρόβλημα αναγνώρισης μικρού μήκους οπτικών χαρακτήρων για πολλές εφαρμογές.

Η αναγνώριση των αριθμών οδών, αποτελεί άλλο ένα ενθαρρυντικό παράδειγμα, το οποίο προσπαθεί να πείσει ότι η μηχανή πλησιάζει τον άνθρωπο στην ικανότητα της όρασης. Ένα παράδειγμα, το οποίο μας δίνει βάσει στο ότι ακόμα ο άνθρωπος είναι καλύτερος από τη μηχανή στην αναγνώριση εικόνων, είναι οι αντιφατικές εικόνες (adversarial images). Οι εικόνες αυτές, αποτελούν προσεκτικά τροποποιημένες

εκδόσεις των αρχικών εικόνων, με τέτοιο τρόπο ώστε στο ανθρώπινο μάτι να μην φαίνεται αισθητή (έως μηδαμινή) διαφορά, αλλά παρ' όλα αυτά να μπερδεύουν τα νευρωνικά δίκτυα σε μεγάλο βαθμό οδηγώντας τα σε τραγικά λάθη [14].



Αντιφατικές Εικόνες: Στην αριστερή στήλη φαίνονται οι αρχικές εικόνες, στη δεξιά στήλη φαίνονται οι αντιφατικές εκδόσεις των αρχικών εικόνων και στη μεσαία στήλη η ανά εικονοστοιχείο διαφορά τους (Εικ. 2.3.3.1)

3. Υλοποίηση

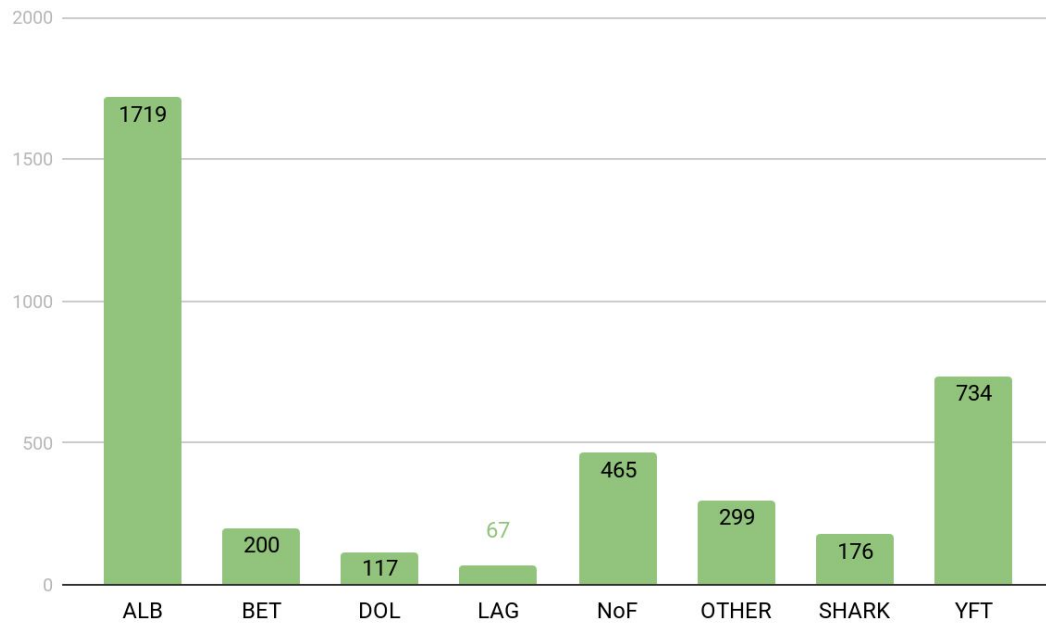
Όπως αναφέρθηκε και στην εισαγωγή, στόχος της εργασίας ήταν η αξιολόγηση εναλλακτικών τεχνικών μηχανικής μάθησης σε δεδομένα εικόνων από το διαγωνισμό “Natural Conservancy Fisheries Monitoring” του διαδικτυακού ιστοτόπου Kaggle. Στην ενότητα αυτή αναπτύσσονται, υλοποιούνται μοντέλα κατηγοριοποίησης εικόνων βασισμένα σε μεθόδους κλασσικής μηχανικής μάθησης και σε τεχνικές ρηχών και πολυεπίπεδων αρχιτεκτονικών νευρωνικών δικτύων. Περιγράφεται η λειτουργία και οι παράμετροι του κάθε μοντέλου, καθώς και τα στάδια προεπεξεργασίας, που υπόκεινται τα δεδομένα πρώτου δοθούν ως είσοδο στην διαδικασία εκπαίδευσης.

3.1. Περιγραφή των Δεδομένων

Τα δεδομένα εκπαίδευσης αποτελούνται από 3777 εικόνες, από 8 κατηγορίες με ονόματα που αντιστοιχούν σε διαφορετικούς τύπους ψαριών ή και εικόνες που δεν εμφανίζεται κανένα ψάρι:

- ALB: Albacore Tuna (*Thunnus alalunga*)
- BET: Bigeye tuna (*Thunnus obesus*)
- DOL: Dolphin fish, Mahi Mahi (*Coryphaena hippurus*)
- LAG: Opah, Moonfish (*Lampris guttatus*)
- NoF: No Fish
- Other: Other fish
- SHARK: Various: Silky, Shortfin Mako
- YFT: Yellowfin tuna (*Thunnus albacares*)

Η κατανομή των κατηγοριών έχει ως εξής:



Κατανομή των εικόνων ανά κατηγορία (Σχ. 3.1.1.1)

3.2: Χρησιμοποιώντας Κλασσική Μηχανική Μάθηση

Σε μια πρώτη βασική προσέγγιση, χρησιμοποιήθηκαν μέθοδοι κλασσικής Μηχανικής Μάθησης, προκειμένου να γίνει η κατηγοριοποίηση των εικόνων. Για τον σκοπό αυτό, χρησιμοποιήθηκε η σουίτα λογισμικού μηχανικής μάθησης WEKA.

3.2.1. Προεπεξεργασία Δεδομένων

Προκειμένου να δοθούν ως είσοδο οι εικόνες στις κλασσικές μεθόδους Μηχανικής Μάθησης που θα χρησιμοποιηθούν, θα πρέπει πρώτα να μετατραπούν οι εικόνες σε γνωρίσματα (features). Αυτό το επιτυγχάνεται στο WEKA μέσω του φίλτρου imageFilters από τον Michael Mayo, η λήψη του οποίου γίνεται μέσω του Package Manager του Weka.

Το φίλτρο αυτό παρέχει διάφορες μεθόδους, με τις οποίες μπορούν να εξαχθούν διάφορα γνωρίσματα για την εικόνα. Εφαρμόζεται πάνω σε ένα γνώρισμα τύπου string το οποίο αποτελεί το όνομα αρχείου εικόνας, προκειμένου να την φορτώσει και να την επεξεργαστεί. Έστερα καταγράφει τα αποτελέσματα που εξήγαγε από το σύνολο εικόνων σε νέα γνωρίσματα στο σύνολο δεδομένων.

Από τα φίλτρα αυτά, επιλέχθηκε το ColorLayoutFilter, το οποίο αποτελεί έναν batch-filter, για την εξαγωγή γνωρισμάτων διατάξεων χρώματος από εικόνες. Η διαδικασία που ακολουθεί είναι να διαιρεί την εικόνα σε 64 μπλοκ, υπολογίζοντας τη μέση τιμή χρώματος για κάθε μπλοκ και δημιουργώντας γνωρίσματα για κάθε μία από τις μέσες τιμές. Αποτελεί ένα από τα πιο γρήγορα φίλτρα στην εφαρμογή τους και από τα πιο αποτελεσματικά.

Η διαδικασία συλλογής των ονομάτων αρχείων των εικόνων και η αντιστοίχιση κάθε ονόματος σε μία ετικέτα, χρειάστηκε ανάπτυξη ξεχωριστού κώδικα, καθώς το WEKA, δεν υποστηρίζει τη λειτουργία αυτή. Ως γλώσσα προγραμματισμού επιλέχθηκε η Python. Το script το οποίο αναλαμβάνει να διαβάσει δοθέντος έναν

κατάλογο τις εικόνες που έχει μέσα και να τις τοποθετήσει ετικέτα, ανάλογα με τους υπο-φακέλους που έχουν οργανωθεί, ονομάζεται `imgtocsv.py`.

Η χρήση του έχει ως εξής:

```
imgtocsv.py (train | test) img_dir [output_dir]
```

Όπου `train` δίνουμε για την μετατροπή του συνόλου εκπαίδευσης σε αρχείο CSV και `test` την μετατροπή του συνόλου δοκιμής σε αρχείο CSV, το οποίο αποθηκεύεται στον φάκελο `output_dir`, αν έχει οριστεί, αλλιώς αποθηκεύεται στον τρέχον κατάλογο. Αυτό το script όπως και τα υπόλοιπα script που χρησιμοποιήθηκαν στα πλαίσια αυτής της εργασίας, βρίσκονται στον ιστότοπο: <https://github.com/panosgemmos/nnimg>.

Σε αυτό το στάδιο, έπειτα από την εφαρμογή του φίλτρου `ColorLayoutFilter`, έχουμε 64 γνωρίσματα και ένα γνώρισμα κλάσης το οποίο υποδεικνύει την κατηγορία της εικόνας.

3.2.2. Αλγόριθμοι

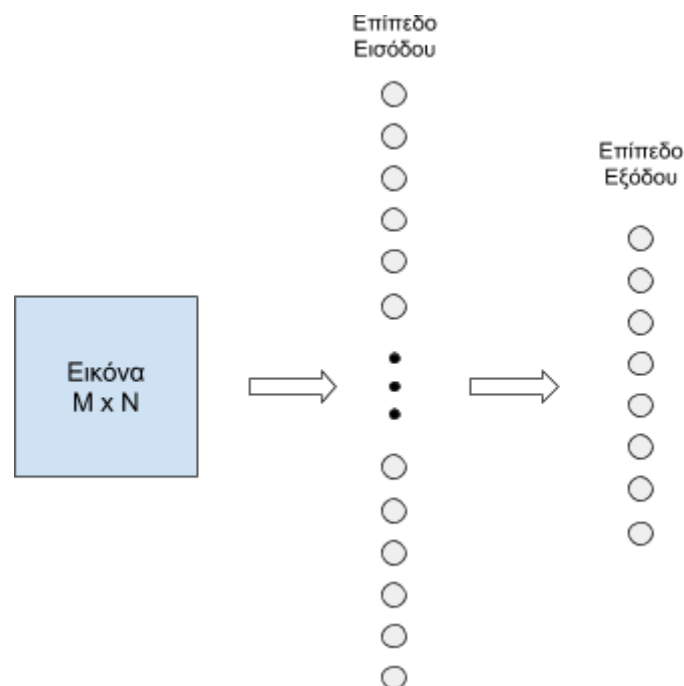
Οι αλγόριθμοι των μοντέλων εκπαίδευσης που χρησιμοποιήθηκαν στα γνωρίσματα που εξήχθησαν κατά το προηγούμενο στάδιο (στάδιο προεπεξεργασίας), είναι οι ZeroR, OneR, Logistic, NaiveBayes, LibSVM, J48, Hoeffding Tree, RandomTree, RandomForest, IBk και KStar. Οι ZeroR και OneR, αποτελούν αλγόριθμους κανόνων (rule-based algorithms), ο Logistic είναι αλγόριθμος Logistic Regression, ο NaiveBayes είναι Μπαϋσιανός αλγόριθμος, ο LibSVM είναι Support Vector Machine αλγόριθμος, οι J48, Hoeffding Tree, RandomTree και RandomForest είναι αλγόριθμοι δέντρων και οι IBk και KStar, είναι lazy αλγόριθμοι. Όλα τα μοντέλα εκπαιδεύονται χρησιμοποιώντας 10-fold Cross Validation. Η μέτρηση της απόδοσης γίνεται μέσω της μετρικής της ακρίβειας, δηλαδή του ποσοστού επιτυχώς κατηγοριοποιημένων εικόνων.

3.3: Χρησιμοποιώντας Ρηγά Νευρωνικά Δίκτυα

Για αυτή την εργασία, χρησιμοποιήθηκαν δύο αρχιτεκτονικές ρηγών νευρωνικών δικτύων. Η BASE_0 και η BASE_1. Η πρώτη αποτελείται από δύο μόνο επίπεδα, το επίπεδο εισόδου και το επίπεδο εξόδου. Η δεύτερη αποτελείται από τρία επίπεδα, ένα εισόδου, ένα κρυφό και ένα εξόδου. Το πλήθος των νευρώνων εισόδου εξαρτάται από το μέγεθος της εικόνας. Για μία μονοχρωματική εικόνα μεγέθους $M \times N$, το μέγεθος εισόδου είναι διάνυσμα μήκους $L = M \times N$, ενώ για μία χρωματική εικόνα το μήκος του διανύσματος είναι ίσο με $L = M \times N \times c$, όπου c το πλήθος των χρωμάτων. Το μέγεθος του επιπέδου εξόδου είναι ίσο με τον αριθμό των κατηγοριών του συνόλου δεδομένων (στην περίπτωση μας 8). Το μέγεθος του κρυφού επιπέδου, στην περίπτωση του μοντέλου BASE_1, αποτελεί υπερ-παράμετρο του μοντέλου.

3.3.1. Αρχιτεκτονική Δικτύου BASE_0

Στο παρακάτω σχήμα φαίνεται η αρχιτεκτονική του ρηγού μοντέλου νευρωνικών δικτύων BASE_0.

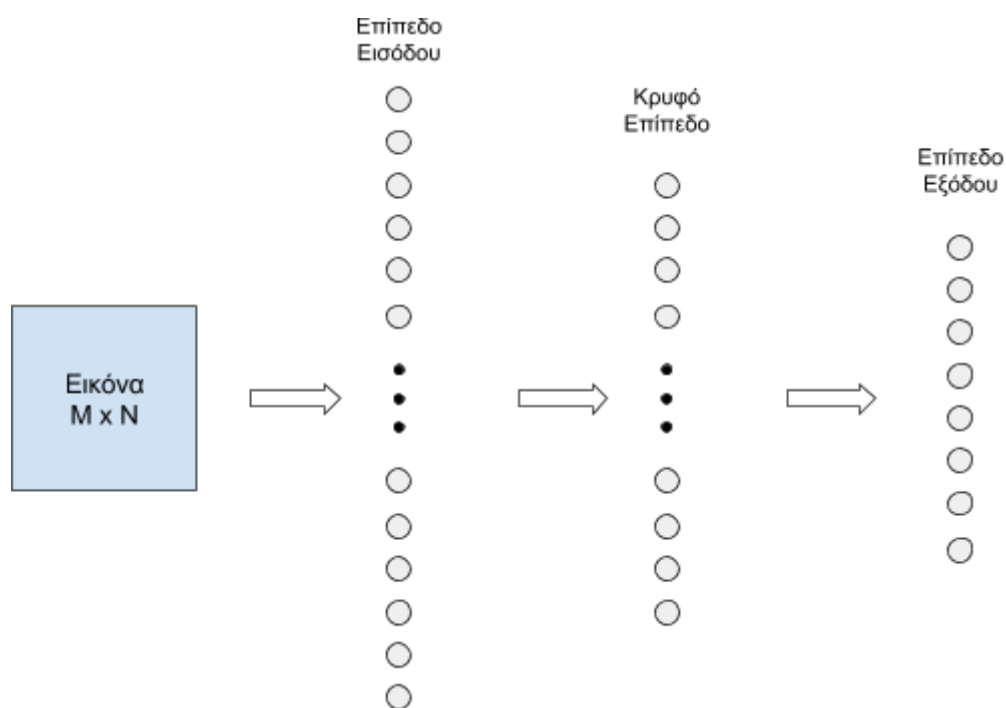


Αρχιτεκτονική Μοντέλου BASE_0 (Σχ. 3.3.1.2)

Πρόκειται για ρηχό νευρωνικό δίκτυο, αποτελούμενο από δύο πλήρως συνδεδεμένα επίπεδα. Το πρώτο αποτελείται από το επίπεδο εισόδου, όπου για κάθε νευρώνα, εφαρμόζεται η σιγμοειδής συνάρτηση στην έξοδό του. Το επίπεδο εξόδου, αποτελεί το επίπεδο softmax, όπου σε κάθε νευρώνα εφαρμόζεται η συνάρτηση softmax. Το τελευταίο επίπεδο, στην υλοποίηση σε κώδικα, αναπαριστάται ως διάνυσμα, οι τιμές του οποίου είναι πιθανότητες που αθροίζουν στο 1. Επομένως, η τιμή 1 σε μία θέση x του διανύσματος, συμβολίζει την ύπαρξη της ετικέτας που αντιστοιχεί στην θέση x . Αντιθέτως το 0 δηλώνει την απουσία ετικέτας ενώ μία τιμή μεταξύ 0 και 1 συμβολίζει την πιθανότητα η συγκεκριμένη εικόνα να αντιστοιχεί στην ετικέτα x .

3.3.2. Αρχιτεκτονική Δικτύου BASE_1

Στο παρακάτω σχήμα φαίνεται η αρχιτεκτονική του ρηχού μοντέλου νευρωνικών δικτύων BASE_1.



Αρχιτεκτονική Μοντέλου BASE_1 (Σχ. 3.3.2.3)

Το μοντέλο BASE_1, έχει ίδια αρχιτεκτονική με το μοντέλο BASE_0, με τη διαφορά ότι ενδιάμεσα των επιπέδων εισόδου και εξόδου, υπάρχει ένα κρυφό πλήρως

συνδεδεμένο επίπεδο νευρώνων. Σε αυτό το επίπεδο, εφαρμόζουμε όπως και στο επίπεδο εισόδου, την σιγμοειδή συνάρτηση σε κάθε ένα από τους νευρώνες.

3.3.3. Προεπεξεργασία Δεδομένων

Η είσοδος των ρηχών νευρωνικών δικτύων, δέχεται για κάθε εικόνα ένα διάνυσμα από τιμές που αντιστοιχούν στις εντάσεις των εικονοστοιχείων της εικόνας. Αυτό σημαίνει ότι θα πρέπει να μετατραπούν οι εντάσεις των εικόνων που είναι σε μορφή πίνακα δύο διαστάσεων σε μία διάσταση. Το σύνολο δεδομένων, αποτελείται από χρωματικές εικόνες τριών καναλιών κόκκινου, πράσινου και μπλέ (RGB). Επομένως υπάρχουν τρεις δισδιάστατοι πίνακες οι οποίοι πρέπει να συγχωνευθούν σε ένα διάνυσμα. Η μέθοδος που ακολουθήθηκε για να επιτευχθεί αυτό, είναι να μετατραπούν ένα-ένα τα κανάλια RGB της εικόνας σε διάνυσμα, λαμβάνοντας έτσι τρία διανύσματα. Συνενώνοντας αυτά τα διανύσματα, λαμβάνουμε ένα διάνυσμα όπως απαιτείται από το επίπεδο εισόδου του νευρωνικού δικτύου.

Επίσης, επειδή οι εικόνες εισόδου παρουσιάζουν διαφορετικά μεγέθη, σε ύψος και πλάτος, χρειάστηκε προτού μετατραπούν οι εικόνες από δισδιάστατες σε μονοδιάστατες, να μετασχηματιστούν σε ένα κοινό μέγεθος. Ο μετασχηματισμός που επιλέχθηκε, ήταν η κεντρική αποκοπή, όπου αποκόπτεται συγκεκριμένο μέρος της εικόνας από το κέντρο της. Για να βρεθεί το κοινό μέγεθος εικόνας, πρέπει να βρεθούν οι μικρότερες διαστάσεις πλάτους και ύψους. Αυτές ορίζουν το νέο κοινό μέγεθος, όπου δινόταν ως όρισμα στην συνάρτηση κεντρικής αποκοπής, όπου η συνάρτηση ύστερα εφαρμόζόταν σε όλες τις εικόνες.

Στο επόμενο στάδιο από την αποκοπή σε κοινή διάσταση και τον μετασχηματισμό σε μία διάσταση, μετατρέπονταν οι ετικέτες της κάθε εικόνας σε onehot διανύσματα. Τα onehot διανύσματα, έχουν μέγεθος όσο και το πλήθος των διαφορετικών ετικετών (στην περίπτωσή μας 8) όπου η κάθε τιμή του διανύσματος, αντιστοιχεί σε μία ετικέτα του συνόλου δεδομένων. Τα onehot διανύσματα παρουσιάζουν την τιμή 1 σε μία συγκεκριμένη θέση, ενώ είναι μηδέν στις υπόλοιπες. Η ύπαρξη 1 σε μία συγκεκριμένη θέση x στο onehot διάνυσμα για μία εικόνα, δηλώνει την αντιστοίχιση της ετικέτας που βρίσκεται στην θέση x στην εικόνα.

0	1	2	3	4	5	6	7
ALB	BET	DOL	LAG	NoF	OTHER	SHARK	YFT
0	0	0	0	1	0	0	0

Παράδειγμα onehot διανύσματος, όπου έχει άσσο στην θέση 4. Αυτό δηλώνει ότι η εικόνα έχει ετικέτα τη NoF (Πιν. 3.3.3.1)

Η τελευταία επεξεργασία που υπέστησαν τα δεδομένα, προτού προχωρήσουμε στο στάδιο εκπαίδευσης του μοντέλου, είναι η σμίκρυνση τους σε μικρότερες εικόνες, οι οποίες θα είναι αντιπροσωπευτικές των αρχικών. Για τον λόγο αυτό διεξήχθη πείραμα στο οποίο ελεγχόταν η καμπύλη μάθησης για 8 μεγέθη, τα οποία ήταν ποσοστά 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.5, 0.6 επί των αρχικών εικόνων. Αποδείχθηκε ότι από το ποσοστό 0.2 και άνω, οι καμπύλες μάθησης έχουν όμοια μορφή και επομένως, μπορεί να χρησιμοποιηθεί το μικρότερο ποσοστό εκ αυτών για το στάδιο εκπαίδευσης, δηλαδή το 0.2. Η σμίκρυνση αυτή έχει ως σκοπό την μείωση του χρόνου και του χώρου που χρειάζεται η διαδικασία εκπαίδευσης. Αν χρησιμοποιούταν όλο το σύνολο δεδομένων, θα χρειαζόταν πάνω από 10GB προσωρινής μνήμης.

3.3.4. Εκπαίδευση μοντέλου BASE_0

Για την εκπαίδευση του μοντέλου BASE_0, χρησιμοποιείται η μέθοδος Stochastic Gradient Descent, μαζί με τον αλγόριθμο back-propagation. Ο ρυθμός μάθησης που δίνεται ως είσοδος αρχικά στο μοντέλο, είναι 0.5. Γίνεται χρήση σιγμοειδών νευρώνων, οι οποίοι αρχικοποιούνται με μηδενικές τιμές. Το μοντέλο εκπαιδεύεται για 100 εποχές σε μικρά πακέτα (mini-batches) μεγέθους 10% εκ του συνολικού συνόλου εκπαίδευσης.

3.3.5. Εκπαίδευση Μοντέλου BASE_1

Για την εκπαίδευση του μοντέλου BASE_1, ακολουθείται ίδια μεθοδολογία με το BASE_0, μονάχα που αντί για να γίνει αρχικοποίηση των νευρώνων σε μηδενικά, γίνεται αρχικοποίηση, χρησιμοποιώντας την συνάρτηση `truncated_normal`, της βιβλιοθήκης Tensorflow. Η `truncated_normal`, υπολογίζει μία τιμή μέσω της κανονικής κατανομής. Το χαρακτηριστικό `truncated` (αποκομμένη) σημαίνει ότι επιστρέφει τιμές μόνο αν απέχουν 2 τυπικές αποκλίσεις από τον μέσο. Αν επιλεγεί μία τιμή εκτός αυτού του εύρους, τότε αυτή απορρίπτεται και επαναλαμβάνεται η διαδικασία, μέχρι να επιστραφεί μία τιμή εντός του εύρους.

3.4: Χρησιμοποιώντας Deep Learning

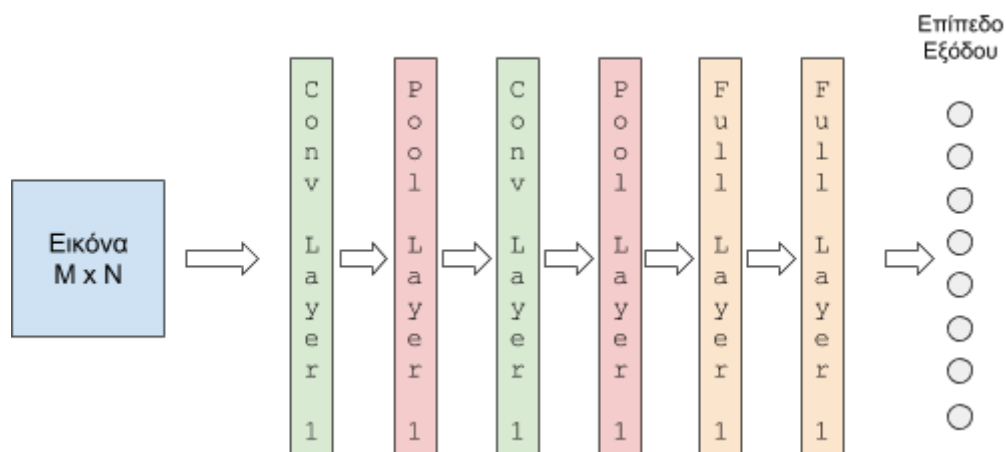
Σε αυτή την ενότητα, κατασκευάζεται ένα συνελκτικό βαθύ νευρωνικό δίκτυο, ονόματι CONV_1, το οποίο αποτελείται από 6 επίπεδα, δύο συνελκτικά επίπεδα, δύο συγκεντρωτικά επίπεδα και δύο πλήρως συνδεδεμένα επίπεδα.

3.4.1 Αρχιτεκτονική του Βαθέως Νευρωνικού Δικτύου CONV_1

Οι νευρώνες που χρησιμοποιούνται είναι τύπου ReLU, καθώς αποδίδουν γενικά καλύτερα έναντι άλλων τύπων νευρώνων στα συνελκτικά νευρωνικά δίκτυα. Για να αποφευχθούν οι “νεκροί νευρώνες”, αρχικοποιούμε τους νευρώνες με μία σχετικά μικρή τυχαία θετική τιμή, την οποία υπολογίζουμε μέσω της συνάρτησης `truncated_normal`, της βιβλιοθήκης Tensorflow, δίνοντας ως παράμετρο τυπική απόκλιση 0.1 .

Όσον αφορά για την ρύθμιση των συνελκτικών επιπέδων, το πρώτο χρησιμοποιεί 32 φίλτρα μεγέθους 5x5 με μέγεθος stride 1 και padding 0, εξάγοντας 32 χαρακτηριστικά. Το δεύτερο είναι ίδιο με το πρώτο, μονάχα που αντί για 32 φίλτρα, χρησιμοποιεί 64 φίλτρα, εξάγοντας 64 χαρακτηριστικά. Οι εξοδοι των συνελκτικών επιπέδων, οδηγούνται σε συγκεντρωτικά επίπεδα, που χρησιμοποιούν φίλτρα μεγέθους 2x2, στα οποία εφαρμόζεται η συνάρτηση μεγίστου.

Στα τελευταία επίπεδα, χρησιμοποιούνται δύο πλήρως συνδεδεμένα επίπεδα. Το πρώτο έχει πλήθος νευρώνων όσα είναι τα κανάλια που εξάγει το αμέσως προηγούμενο συγκεντρωτικό πεδίο, επί το μέγεθος του κάθε καναλιού. Δεν μπορούμε να γνωρίζουμε εξ αρχής το μέγεθος του πλήρους συνδεδεμένου επιπέδου στο συγκεκριμένο dataset, καθώς εξαρτάται κάθε φορά από το πως θα χωριστεί το σύνολο εκπαίδευσης και το σύνολο επαλήθευσης. Η εξάρτηση αυτή οφείλεται στο ότι οι εικόνες, παρουσιάζουν διαφορετικά μεγέθη μέσα στα δεδομένα. Έτσι, μπορεί κατά μία διαδικασία τυχαίας επιλογής των συνόλων εκπαίδευσης και επαλήθευσης να βρίσκονται διαφορετικού μεγέθους εικόνες στο σύνολο εκπαίδευσης, κάτι το οποίο σημαίνει ότι ίσως να υπολογιστεί, διαφορετικό κοινό ελάχιστο μέγεθος. Το δεύτερο, έχει 1024 νευρώνες. Για τη μείωση της υπερ-εκπαίδευσης, χρησιμοποιείται η μέθοδος dropout, η οποία εφαρμόζεται μεταξύ του πρώτου και του δεύτερου πλήρους συνδεδεμένου επιπέδου. Το επίπεδο εξόδου, αποτελείται από επίπεδο softmax.



Αρχιτεκτονική Μοντέλου CONV_1 (Σχ. 3.4.1.1)

3.4.2 Προεπεξεργασία Δεδομένων

Η προεπεξεργασία που χρειάστηκαν τα δεδομένα προκειμένου να δοθούν ως είσοδος στο νευρωνικό δίκτυο, είναι μόνο η αποκοπή των εικόνων ώστε να έχουν το ίδιο μέγεθος. Δεν χρειάζεται σε αυτή τη περίπτωση να μετατρέψουμε τις τριών καναλιών χρώματος δισδιάστατες εικόνες σε διάνυσμα μίας διάστασης, καθώς τα συνελκτικά νευρωνικά δίκτυα δέχονται πολλαπλών διαστάσεων δεδομένα.

3.4.3 Εκπαίδευση Μοντέλου

Για την εκπαίδευση του συνελκτικού μοντέλου νευρωνικών δικτύων, χρειάζεται να οριστούν μερικές παράμετροι μάθησης και βελτιστοποίησης. Η συνάρτηση απωλειών, είναι η cross-entropy, η οποία πετυχαίνει συνήθως πολύ καλύτερα αποτελέσματα στα νευρωνικά δίκτυα, από ότι η κλασσική μέθοδος ελαχίστου τετραγώνων. Επίσης, αντί για την σύννηθες μέθοδο Gradient Descent, θα χρησιμοποιήσουμε μια πιο μοντέρνα μέθοδο, την ADAM, ως συνάρτηση βελτιστοποίησης της συνάρτησης απωλειών. Ως ρυθμό μάθησης (learning reate), χρησιμοποιούμε την τιμή 0.001. Οι παράμετροι της παραπάνω αρχιτεκτονικής, καθώς και οι διασυνδέσεις των νευρώνων, εμπνεύστηκαν από μία αρχιτεκτονική που χρησιμοποιείται στην ιστοσελίδα του Tensorflow, με σκοπό την κατηγοριοποίηση εικόνων [15]. Οι τιμές που ορίζονται για τις παραμέτρους, αποτελούν προκαθορισμένες τιμές οι οποίες εμπειρικά λειτουργούν καλά για τη πλειοψηφία των προβλημάτων κατηγοριοποίησης εικόνας. Σε κάθε περίπτωση όμως, θα πρέπει σε μελλοντική βελτίωση αυτής της εργασίας, να δοκιμαστούν περισσότερες τιμές για τις παραμέτρους αυτές.

4. Αποτελέσματα

4.1. Κλασική μηχανική μάθηση

4.1.1 Αξιολόγηση στα Δεδομένα Εκπαίδευσης

Ως αρχικά μοντέλα εκπαίδευσης, χρησιμοποιήθηκαν αυτά με τους πιο απλούς αλγόριθμους από όλους, δηλαδή οι rule-based, έχοντάς τους ως ένα κατώτατο κατώφλι για την απόδοση των υπόλοιπων μοντέλων. Ο ZeroR, αυτό που κάνει είναι να προβλέπει συνέχεια την κλάση η οποία στο σύνολο δεδομένων είχε τη μεγαλύτερη συχνότητα εμφάνισης. Η ακρίβεια που πέτυχε ήταν 45.5123%. Ο OneR, οποίος είναι rule based αλγόριθμος, αποτελείται από κανόνες if-then-else οι οποίοι εφαρμόζονται στο γνώρισμα το οποίο διασπά καλύτερα τους κόμβους. Η ακρίβεια που πέτυχε ήταν 50.3574%.

Το μοντέλο το οποίο κατασκευάστηκε χρησιμοποιώντας τον αλγόριθμο Logistic, απέδωσε ακρίβεια 62.9865%. Ποσοστό μεγαλύτερο από το προηγούμενο κατώφλι. Το αποτέλεσμα της ακρίβειας του NaiveBayes ήταν 49.5367%, χειρότερα και από το κατώφλι του OneR.

Στο τέλος απομένουν οι αλγόριθμοι που πέτυχαν τις μεγαλύτερες αποδόσεις. Οι αλγόριθμοι δέντρων τα πήγαν αρκετά καλά, με ποσοστά ακρίβειας 86.4972% για τον J48 και τον RandomTree και 94.8636% για τον RandomForest. Αν και η πλειοψηφία των δενδρικών αλγορίθμων είχε υψηλή ακρίβεια, ο αλγόριθμος Hoeffding Tree, πέτυχε μόλις 49.1131%. Δηλαδή χειρότερα και από το κατώφλι που ορίζει ο OneR. Ο LibSVM κατέφερε ποσοστό ακρίβειας 92.6397%, φτάνοντας σε επίπεδα τον RandomForest. Οι καλύτεροι βέβαια όλων αναδείχθηκαν οι lazy αλγόριθμοι, πετυχαίνοντας ακρίβεια 95.4461% ο IBk και 95.7109% ο KStar.

Νικητής όμως αναδεικνύεται ο IBk, καθώς χρειάστηκε τη λιγότερη ώρα να εκπαιδευθεί και να επαληθευθεί σε αντίθεση με τα υπόλοιπα μοντέλα, τα οποία χρειάστηκαν αρκετό χρόνο για την εκπαίδευση αλλά και για την επαλήθευση. Επίσης,

νικά όλα τα μοντέλα σε ακρίβεια, πλην του KStar, με διαφορά περίπου 0.3%. Επειδή ο KStar όμως ήταν αρκετά πιο αργός, απορρίφθηκε ως ένα από τα μοντέλα από τα οποία θα συνεχίσουν στο επόμενο στάδιο του fine-tuning.

Πειράζοντας τις παραμέτρους του IBk, καμία δεν έφερε καλύτερο αποτέλεσμα από ότι οι προκαθορισμένες παράμετροι, πέραν της αλλαγής της παραμέτρου απόστασης από Ευκλείδεια σε Manhattan. Η ακρίβεια έπειτα από αυτή την αλλαγή, αυξήθηκε στο 95.8186%. Καλύτερα και από τον KStar.

Επομένως, η καλύτερη μέθοδος κλασσικής μηχανικής μάθησης για αυτά τα δεδομένα, είναι το μοντέλο IBk. Οι απόδοση αυτή όμως, αφορά τα δεδομένα εκπαίδευσης, δηλαδή σύνολο εκπαίδευσης και σύνολο επαλήθευσης. Πόσο καλά αποδίδει το μοντέλο αυτό, σε άγνωστα δεδομένα (σύνολο δοκιμής);

4.1.2. Αξιολόγηση στα Δεδομένα Ελέγχου

Συνήθως, οι διαγωνισμοί στο Kaggle, δεν δίνουν τις ετικέτες του συνόλου δοκιμής στο κοινό σε διαγωνισμούς με χρηματικό έπαθλο όπως ο “The Nature Conservancy Fisheries Monitoring”. Ο μόνος τρόπος να δοκιμάσει κάποιος το μοντέλο του είναι να ανεβάσει τις προβλέψεις του στο Kaggle, το οποίο θα του υπολογίσει την απόδοση που είχαν οι προβλέψεις πάνω στο σύνολο δοκιμής. Για την δημιουργία του αρχείου το οποίο περιέχει τις προβλέψεις, χρειάστηκε χρήση εξωτερικού κώδικα. Το script `wekatokaggle.py`, μετατρέπει τις προβλέψεις που εξάγει το WEKA, στη μορφή που χρειάζεται η ιστοσελίδα Kaggle, για να τις αξιολογήσει.

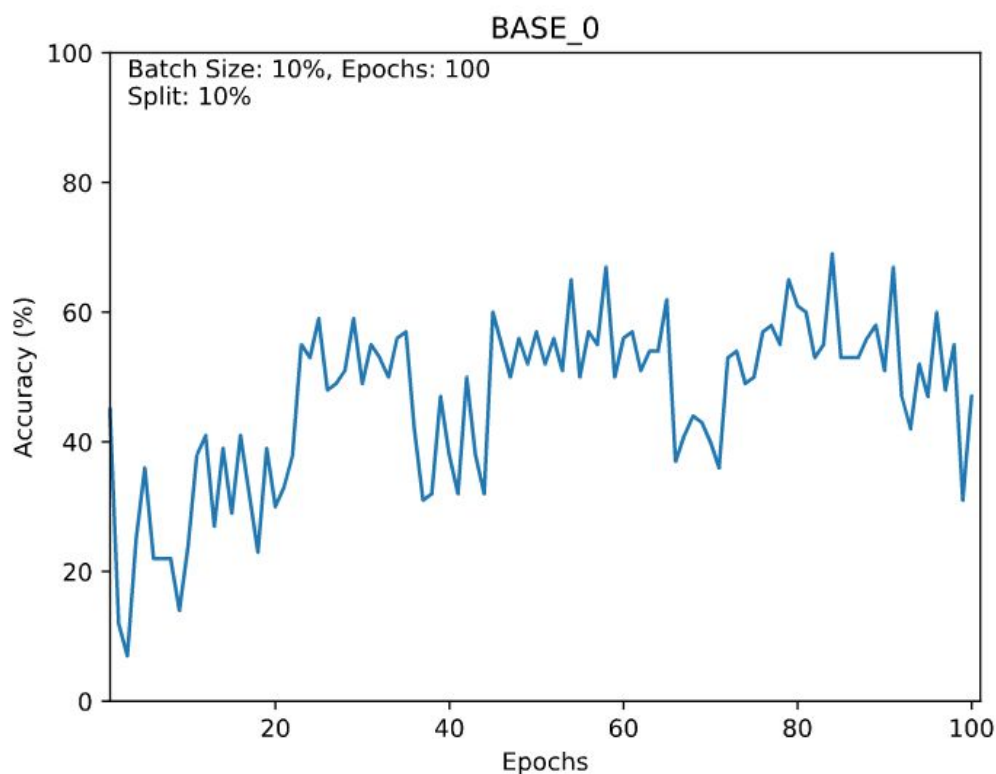
Δυστυχώς όμως, το μοντέλο IBk, απέδωσε πολύ άσχημα. Αυτό σημαίνει ότι είχε υπερ-εκπαιδευτεί στο σύνολο εκπαίδευσης και για το λόγο αυτό εμφάνιζε τόσο υψηλά ποσοστά στο σύνολο επαλήθευσης. Το φαινόμενο αυτό οφείλεται στο ότι το σύνολο εκπαίδευσης ήταν αρκετά διαφορετικό από το σύνολο δοκιμής, όσον αφορά το υπόβαθρο των εικόνων, το οποίο σε συνδυασμό με το φίλτρο εικόνας και τον τρόπο ταξινόμησης, έμαθε το μοντέλο να κατηγοριοποιεί βάσει υποβάθρου και όχι των ψαριών. Έτσι, επειδή το κάθε είδος ψαριών στο σύνολο εκπαίδευσης βρισκόταν σε συγκεκριμένο υπόβαθρο, ο IBk, ήταν καλός στο να βρίσκει το συγκεκριμένο υπόβαθρο και άρα το ταξινομούσε σωστά. Στο σύνολο όμως δοκιμής, όπου το υπόβαθρο είναι διαφορετικό από αυτό του συνόλου εκπαίδευσης, ο IBk αποδίδει με τεράστια αποτυχία να βρει ένα υπόβαθρο που να μοιάζει με αυτά που “γνωρίζει” ήδη.

4.2: Χρησιμοποιώντας Ρηγά Νευρωνικά Δίκτυα

Σε αυτό το κεφάλαιο, παρουσιάζονται βασικές, απλοί μέθοδοι νευρωνικών δικτύων, χρησιμοποιώντας ρηχές αρχιτεκτονικές. Η απόδοσή τους συγκρίνεται με αυτή των κλασσικών μεθόδων Μηχανικής Μάθησης, όπου αποδεικνύεται ότι αποδίδουν κατά πολύ καλύτερα.

4.2.1 Αποτελέσματα στα Δεδομένα Εκπαίδευσης

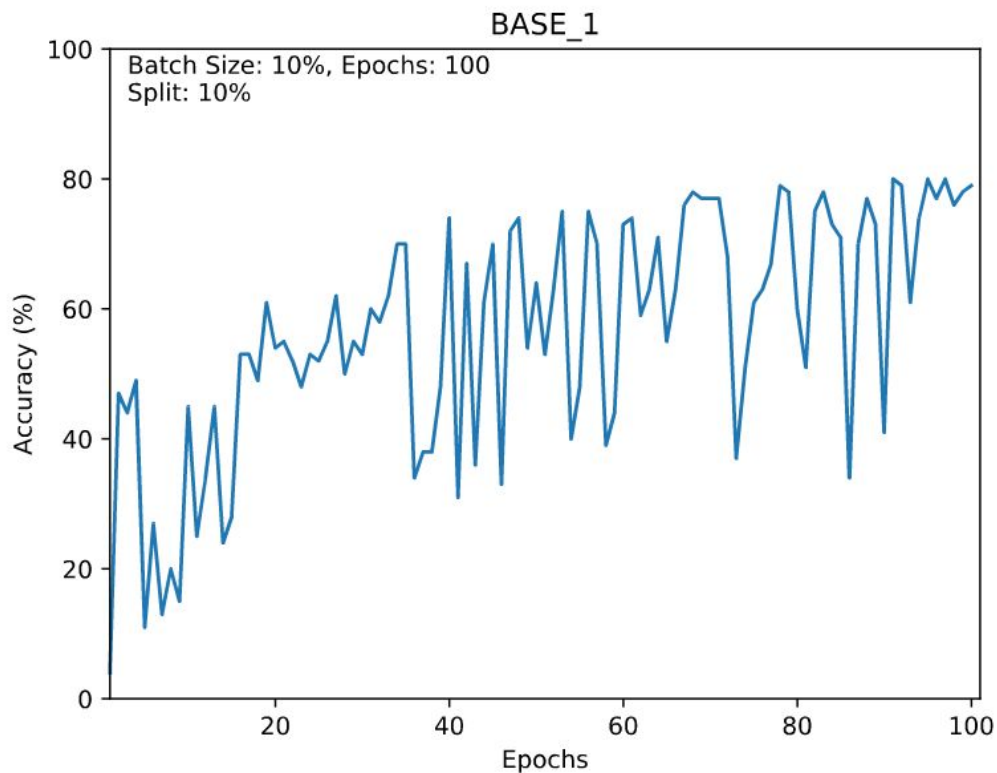
Εκπαιδεύοντας το ρηχό νευρωνικό δίκτυο BASE_0, στο 90% του συνόλου δεδομένων, κρατώντας το 10% για επαλήθευση, λαμβάνουμε την εξής καμπύλη μάθησης:



Μοντέλο BASE_0 - Καμπύλη Μάθησης (Σχ. 4.2.1.1)

Όπως παρατηρούμε από τη παραπάνω καμπύλη, το νευρωνικό δίκτυο δε βελτίωσε την απόδοσή του όσο προχωρούσε η διαδικασία εκπαίδευσης (epochs) και θα μπορούσαμε να ισχυριστούμε ότι το δίκτυό μας τα πήγε λίγο καλύτερα από το να μάντευε στην τύχη.

Για το BASE_1, εκπαιδεύοντας με τις ίδιες παραμέτρους με το BASE_0 με πλήθος νευρώνων κρυφού επιπέδου στα 100, έχουμε την παρακάτω καμπύλη μάθησης.



Μοντέλο BASE_1 - Καμπύλη Μάθησης (Σχ. 4.2.1.2)

Παρατηρούμε ότι το μοντέλο BASE_1, αποδίδει καλύτερα από ότι το BASE_0. Αυτό οφείλεται στο ότι περιέχει ένα ακόμη παραπάνω επίπεδο, το οποίο του επιτρέπει να συλλέξει ακόμη περισσότερα στοιχεία από ότι το BASE_0, για το ποια είναι η σωστή κατηγοριοποίηση μιας εικόνας. Αυτό έχει ως αποτέλεσμα με το πέρασμα των κύκλων εκπαίδευσης (epochs) να βελτιώνεται σταδιακά η απόδοση του μοντέλου μας. Σε κάθε περίπτωση όμως θα πρέπει να ελέγξουμε και το ενδεχόμενο της υπερκπαίδευσης του μοντέλου μας στα δεδομένα εκπαίδευσης (overfitting) και για το σκοπό αυτό θα χρειαστεί να το αξιολογήσουμε στα δεδομένα ελέγχου.

4.2.2 Αποτελέσματα στα Δεδομένα Ελέγχου

Ανεβάζοντας τις προβλέψεις για το σύνολο δοκιμής στην ιστοσελίδα του διαγωνισμού “Nature Conservancy Fisheries Monitoring” στο Kaggle, από το μοντέλο η BASE_0, επετεύχθη score 2.842, το οποίο αντιστοιχεί κοντά στην 360η θέση της κατάταξης. Το score, μετριέται μέσω της συνάρτησης multi-class logarithmic loss, η οποία ορίζεται ως εξής:

$$logloss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$

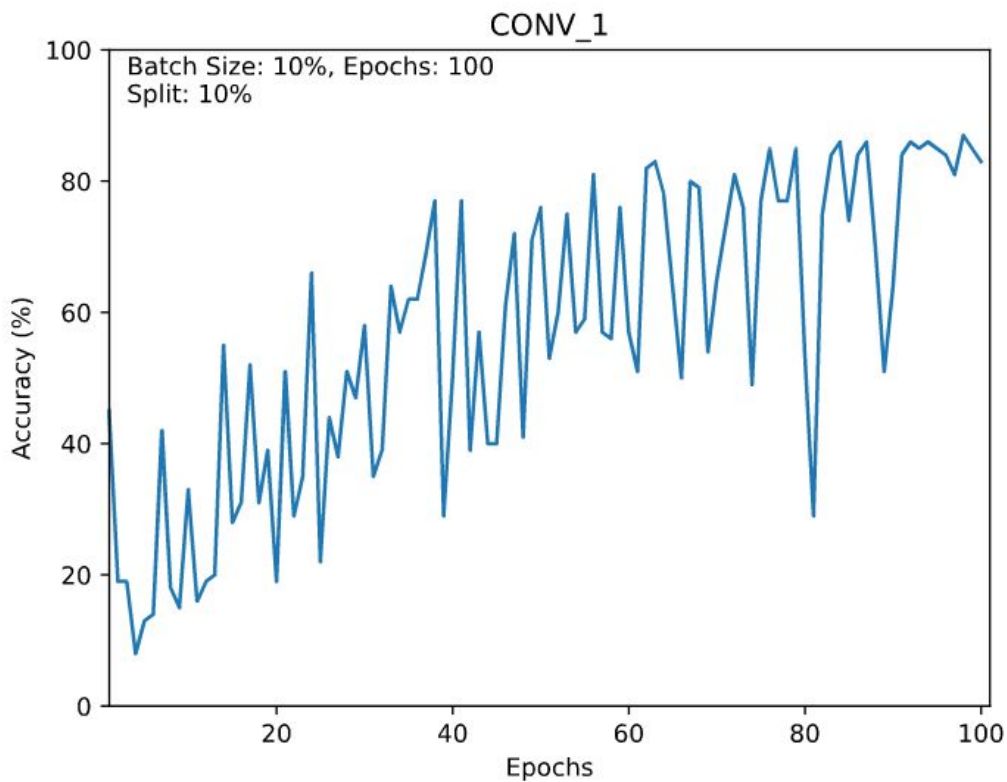
όπου N είναι το πλήθος των εικόνων στο σύνολο δοκιμής, M είναι το πλήθος των ετικετών, log ο φυσικός λογάριθμος, y_{ij} ισούται με 1 αν η παρατήρηση i ανήκει στην κλάση j και 0 αλλιώς, και p_{ij} είναι η προβλεπόμενη πιθανότητα ότι μία παρατήρηση j ανήκει στην κλάση i. Το δίκτυο αυτό, αν και ρηχό, τα πήγε πολύ καλύτερα από τις κλασσικές μεθόδους Μηχανικής Μάθησης, οι οποίες στο συγκεκριμένο πρόβλημα, απέτυχαν παταγωδώς. Το μοντέλο BASE_1, πέτυχε score 2.627, το οποίο στην κατάταξη αντιστοιχεί περίπου στην 350η θέση. Αυτό επαληθεύει το γεγονός ότι το μοντέλο BASE_1 υπερτερεί του BASE_0 σε απόδοση.

4.3. Χρησιμοποιώντας Deep Learning

Σε αυτό το κεφάλαιο παρουσιάζονται αποτελέσματα πολυεπίπεδων αρχιτεκτονικών νευρωνικών δικτύων Deep Learning, όπου αποδεικνύεται ότι αποδίδουν πολύ καλύτερα σε σχέση με τις ρηχές αρχιτεκτονικές και κατ’ επέκταση με τις κλασσικές τεχνικές Μηχανικής Μάθησης.

4.3.1 Αποτελέσματα στα Δεδομένα Εκπαίδευσης

Εκπαιδεύοντας το βαθύ πολυεπίπεδο συνελκτικό νευρωνικό δίκτυο CONV_1, στο 90% του συνόλου δεδομένων, κρατώντας το 10% για επαλήθευση, λαμβάνουμε την εξής καμπύλη μάθησης:



Μοντέλο CONV_1 (Σχ. 4.3.1.1)

Όπως παρατηρούμε από τη παραπάνω καμπύλη, το νευρωνικό δίκτυο απέδωσε καλύτερα και από τα δύο προηγούμενα ρηχά νευρωνικά δίκτυα.

4.3.2 Αποτελέσματα στα Δεδομένα Ελέγχου

Ανεβάζοντας τις προβλέψεις για το σύνολο δοκιμής στην ιστοσελίδα του διαγωνισμού “Nature Conservancy Fisheries Monitoring” στο Kaggle, επετεύχθη score 2.342, το οποίο αντιστοιχεί κοντά στην 330η θέση της κατάταξης του διαγωνισμού. Το δίκτυο αυτό, πέτυχε απόδοση πολύ μεγαλύτερη από τα ρηχά νευρωνικά δίκτυα, και κατ’ επέκταση πολύ καλύτερα από τις κλασικές μεθόδους Μηχανικής Μάθησης.

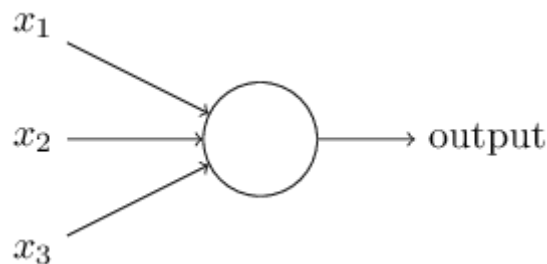
5. Συμπεράσματα - μελλοντικές επεκτάσεις

Από τα πειράματα που εκτελέστηκαν, μπορεί να βγει το συμπέρασμα πως όσα περισσότερα είναι τα επίπεδα των νευρωνικών δικτύων, τόσο καλύτερη είναι η απόδοσή τους. Αρκεί όμως να έχουμε θέσει τις κατάλληλες παραμέτρους, διότι υπάρχει περίπτωση να καταλήξουμε σε χειρότερα αποτελέσματα. Επίσης, η ύπαρξη όχι μόνο πλήρως συνδεδεμένων επιπέδων, αλλά και επιπέδων με άλλα χαρακτηριστικά όπως τα επίπεδα συνέλιξης και συγκέντρωσης, μπορούν να αποδώσουν καλύτερα από ότι αν η αρχιτεκτονική του νευρωνικού δικτύου περιείχε μόνο πλήρως διασυνδεδεμένα επίπεδα. Ένα δεύτερο συμπέρασμα που προέκυψε από την εκτέλεση των πειραμάτων για αυτή την εργασία, είναι ότι αν και με περισσότερα επίπεδα, επιτυγχάνεται μεγαλύτερη απόδοση, χρειάζεται περισσότερος χρόνος για να εκπαιδευτεί το μοντέλο.

Οι μελλοντικές επεκτάσεις που θα μπορούσαν να εφαρμοστούν σε αυτή την εργασία, είναι η περαιτέρω στατιστική ανάλυση των αποτελεσμάτων και ο έλεγχος σημαντικότητας αυτών. Επίσης, θα μπορούσαν να παρουσιαστούν περισσότερες αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων, οι οποίες χρησιμοποιούν ακόμα περισσότερα επίπεδα και περισσότερους νευρώνες. Αυτό δεν έγινε στην τρέχον εργασία, καθώς δεν υπήρχαν οι κατάλληλοι υπολογιστικοί πόροι που χρειάζονται τέτοιου είδους βαθιά νευρωνικά δίκτυα.

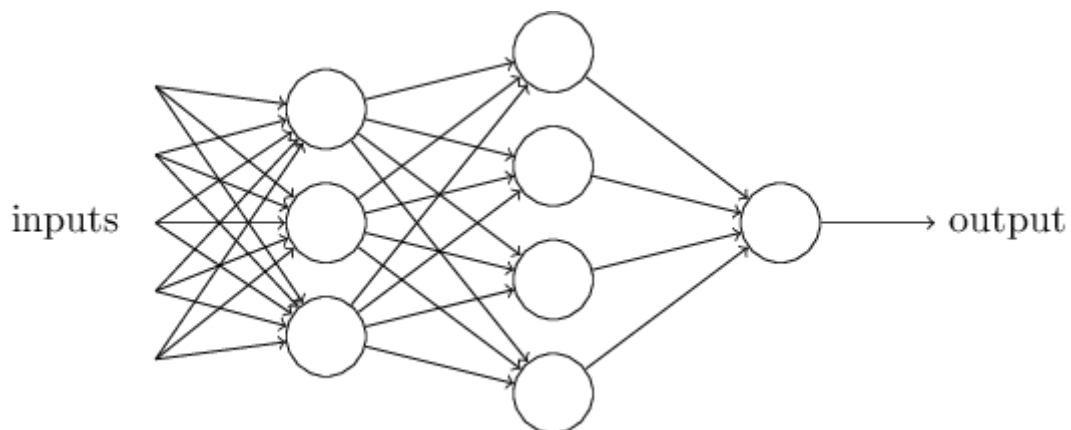
Παράρτημα Α: Αρχιτεκτονική Ρηχών Νευρωνικών Δικτύων

Τα τεχνητά νευρωνικά δίκτυα έχουν εμπνευστεί από την λειτουργία του βιολογικού νευρώνα. Ο νευρώνας δέχεται ως είσοδο σήματα από άλλους νευρώνες, τα οποία συναθροίζει και εξάγει σε άλλους νευρώνες, εφαρμόζοντας ενδιάμεσα μία συνάρτηση ενεργοποίησης (activation function). Παρακάτω φαίνεται μία εικόνα ενός νευρώνα, όπου απεικονίζονται τα πολλά σήματα που εισέρχονται σε αυτόν και η ένα σήμα εξόδου το οποίο υπολογίζει, βάσει τις εισόδους που δέχεται.



Νευρώνας (Εικ. Α.1)

Πολλοί νευρώνες διασυνδεδεμένοι μαζί, δημιουργούν ένα δίκτυο νευρώνων, γνωστό και ως νευρωνικό δίκτυο. Παρακάτω απεικονίζεται ένα απλό δίκτυο νευρώνων:



Δίκτυο από νευρώνες (Εικ. Α.2)

Όπως και ένας μεμονωμένος νευρώνας, έτσι και το δίκτυο νευρώνων έχει είσοδο και έξοδο. Αν και ο υπολογισμός της εξόδου ενός νευρώνα είναι εύκολος, ο υπολογισμός της εξόδου ενός νευρωνικού δικτύου, είναι πιο περίπλοκος, καθώς εξαρτάται από όλα τα προηγούμενα σήματα τα οποία έχουν ληφθεί από τους προηγούμενους νευρώνες.

Στα τεχνητά νευρωνικά δίκτυα, αναθέτουμε για κάθε σύνδεση μεταξύ δύο νευρώνων και από ένα βάρος. Πέρα από τα βάρη, χρησιμοποιείται και άλλη μία μεταβλητή η οποία υπολογίζεται αφού έχουν αθροιστεί όλα τα προηγούμενα σήματα (δηλαδή τα βάρη). Αυτή ονομάζεται κλίση ή bias. Κατά την εκπαίδευση ενός μοντέλου τεχνητών νευρωνικών δικτύων, ο σκοπός είναι να βρεθούν τα κατάλληλα βάρη και κλίσεις, ώστε για άγνωστη είσοδο, και μέσω της γνώσης που συσσωρεύεται σε αυτές τις μεταβλητές, να μπορεί να αποδοθεί η κατάλληλη ετικέτα ή κατηγορία σε αυτή. Προκειμένου να γίνει κάτι τέτοιο, θα πρέπει το μοντέλο να έχει εκπαιδευτεί σε δεδομένα όπου η κατηγορία τους είναι γνωστή, έτσι ώστε να συλλεχθεί πληροφορία μέσω αυτών των δεδομένων η οποία θα είναι χρήσιμη για την κατηγοριοποίηση άγνωστων δεδομένων.

Υπάρχουν πολλοί τρόποι να εκπαιδεύσει κάποιος ένα νευρωνικό δίκτυο. Οι πιο συχνοί από αυτούς είναι μέσω της μεθόδου gradient descent και της μεθόδου backpropagation. Το ζευγάρι αυτό, αποτελεί το βασικό συστατικό στοιχείο εκπαίδευσης αρχιτεκτονικών νευρωνικών δικτύων μέχρι και σήμερα.

Παράρτημα Β: Λογισμικό Υλοποίησης Νευρωνικών Δικτύων

Για την κατασκευή των νευρωνικών δικτύων, χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python, σε συνδυασμό με τις βιβλιοθήκες Numpy, Tensorflow, SciPy, Matplotlib, Scikit-learn και lycon. Επιλέχθηκε η γλώσσα Python, έναντι άλλων γνωστών γλωσσών προγραμματισμού, καθώς περιέχει μεγαλύτερη ευελιξία για εργασίες που αφορούν την Μηχανική Μάθηση και τις επιστημονικές εφαρμογές. Επίσης οι περισσότερες και οι πιο γνωστές βιβλιοθήκες λογισμικού για Μηχανική Μάθηση και κυρίως για νευρωνικά δίκτυα, έχουν συνταχθεί για την γλώσσα Python.

Η βιβλιοθήκη Numpy, προσθέτει στην Python, την υποστήριξη για μεγάλους πολυδιάστατους πίνακες, μαζί με ένα σύνολο υψηλού επιπέδου συναρτήσεων, οι οποίες εφαρμόζονται σε αυτούς. Χρησιμοποιήθηκε για την αποθήκευση των νευρωνικών δικτύων σε πίνακες καθώς και για την εκτέλεση πράξεων και εφαρμογή συναρτήσεων πάνω στους νευρώνες.

Η βιβλιοθήκη Tensorflow, χρησιμοποιήθηκε για την διαχείριση των μοντέλων νευρωνικών δικτύων. Αυτή περιλαμβάνει:

- τις λειτουργίες κατασκευής του μοντέλου, ορίζοντας τις διασυνδέσεις μεταξύ των νευρώνων και δημιουργώντας τα επίπεδά τους.
- Τον ορισμό των παραμέτρων μάθησης (learning parameters).
- Τον ορισμό των υπερ-παραμέτρων του μοντέλου (π.χ μέγεθος δέσμης, αριθμός εποχών).
- Την εκπαίδευση, την επαλήθευση και την δοκιμή του μοντέλου.
- Την αποθήκευση και την φόρτωση του μοντέλου.

Το SciPy, αποτελεί μία σουίτα αλγορίθμων και συναρτήσεων που αφορούν τις επιστήμες και την μηχανική. Στο κώδικα για αυτή την εργασία, χρησιμοποιήθηκαν οι συναρτήσεις εξαγωγής συχνотήτων δεδομένων που παρέχει, με σκοπό την εξαγωγή στατιστικών για τα δεδομένα.

Το Matplotlib, αποτελεί βιβλιοθήκη η οποία παρέχει λειτουργίες για την δημιουργία γραφικών παραστάσεων και διαγραμμάτων. Χρησιμοποιήθηκε για την δημιουργία γραφικών παραστάσεων των αποτελεσμάτων που λήφθηκαν από την διεξαγωγή πειραμάτων στα νευρωνικά δίκτυα (π.χ learning curves).

Η Scikit-learn, είναι μια από τις πιο γνωστές βιβλιοθήκες μηχανικής μάθησης για την γλώσσα προγραμματισμού Python. Από αυτή, χρησιμοποιήθηκαν στον κώδικα της πτυχιακής εργασίας συναρτήσεις που αφορούν την διαχείριση του συνόλου δεδομένων (π.χ διαχωρισμός σε σύνολο εκπαίδευσης και σύνολο επαλήθευσης, διατηρώντας την κατανομή των ετικετών).

Η lychon, είναι μία μικρή βιβλιοθήκη η οποία προσφέρει συναρτήσεις για ταχύτατη αλλαγή μεγέθους των εικόνων καθώς και ταχύτατη αποθήκευση και φόρτωση εικόνων. Ο κύριος λόγος που προστέθηκε στο κώδικα της πτυχιακής εργασίας, είναι για να μειώσει τον χρόνο που χρειάζεται να γίνει η μαζική αλλαγή μεγέθους στις εικόνες που δίνονται ως είσοδο στα μοντέλα των νευρωνικών δικτύων, καθώς και για να μειώσει τον χρόνο φόρτωσης των εικόνων στην μνήμη.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- [2] Werbos, P. J. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. *Doctoral Dissertation, Applied Mathematics, Harvard University, MA*.
- [3] Spall, James C. (2003), Introduction to Stochastic Search and Optimization, Wiley, ISBN 978-0-471-33052-3.
- [4] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [5] Kaggle “Nature Conservancy Fisheries Monitoring” Competition
<https://www.kaggle.com/c/the-nature-conservancy-fisheries-monitoring>.
- [6] Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1), 221-231.
- [7] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 248-255). IEEE.
- [8] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [9] Le, Q. V. (2013, May). Building high-level features using large scale unsupervised learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 8595-8598). IEEE.
- [10] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [11] LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., ... & Vapnik, V. (1995). Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, 261, 276.

- [12] Karpathy, A. (2014). What I learned from competing against a convnet on ImageNet. 2015-01-24]. URL:
<http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/>.
- [13] Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., & Shet, V. (2013). Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*.
- [14] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.
- [15] “Deep MNIST for Experts” URL:
https://www.tensorflow.org/get_started/mnist/pros