

PNEUMONIA DETECTION (Nov2021)

Ραδής Μάριος Τουμπαλίδης

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ,
ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΕΛΛΑΔΟΣ
it185289@it.teithe.gr

ABSTRACT

Τα τελευταία χρόνια το αντικείμενο της μηχανικής μάθησης επεκτείνεται με εκθετικούς βαθμούς. Στο γνωστό διαγωνισμό ImageNet μοντέλο μηχανικής μάθησης κατάφερε να πετύχει ακρίβεια καλύτερη από τον άνθρωπο, επομένως όλο και περισσότεροι ερευνητές αρχίζουν να ασχολούνται με το αντικείμενο. Συγκεκριμένα η τεχνολογία των συνελκτικών νευρωνικών δικτύων έχει φέρει μία “επανάσταση” στο χώρο της αναγνώρισης εικόνων και προτύπων και της υπολογιστικής όρασης. Σε αυτό το άρθρο θα αναλυθούν τα μοντέλα μηχανικής μάθησης που κατασκευάστηκαν για να πραγματοποιούν ταξινόμηση σε ένα dataset εικόνων “x-ray” οι οποίες προβάλλουν θώρακες ατόμων.

1. INTRODUCTION

Το πρόβλημα που έχουμε να αντιμετωπίσουμε αποτελεί πρόβλημα ταξινόμησης εικόνων σε κλάσεις, και έχουμε ένα σύνολο 5840 εικόνων. Σε αυτό το άρθρο θα αναλυθούν τα μοντέλα μηχανικής μάθησης που κατασκευάστηκαν για να πραγματοποιούν ταξινόμηση σε ένα dataset εικόνων “x-ray” οι οποίες προβάλλουν θώρακες ατόμων. Οι εικόνες αυτές κατηγοριοποιούνται σε τρεις κατηγορίες, αυτοί που δεν έχουν πνευμονία και αυτοί που έχουν πνευμονία είτε από ιό είτε από βακτήριο. Το σύνολο των εικόνων εκπαίδευσης είναι 4672 από τις οποίες 1227 είναι υγιή, 2238 έχουν πνευμονία από βακτήρια και 1207 έχουν πνευμονία από ιό ενώ το σύνολο των εικόνων ελέγχου είναι 1168. Για το συγκεκριμένο πρόβλημα ταξινόμησης κατασκευάστηκαν τέσσερα μοντέλα μηχανικής μάθησης καθώς και υλοποιήθηκαν κάποιες διαδικασίες προ-επεξεργασίας των εικόνων.

2. PREPROCESS - PARAMETER TUNING

Σε αυτό το σημείο θα αναφέρουμε πως η προεπεξεργασία των εικόνων ήταν αναγκαία για την σωστή τροφοδότηση των δικτύων. Ο κύριος λόγος ανάγκης προεπεξεργασίας των εικόνων είναι η διαφορά στις διαστάσεις τις κάθε εικόνας. Συγκεκριμένα, για λόγους οικονομίας και

υπολογιστικής δύναμης η διάσταση των εικόνων για εκπαίδευση καθώς και για έλεγχο (δηλαδή η διάσταση των εικόνων στην εισοδο των δικτύων) είναι (244,224). Για την αλλαγή διαστάσεων των εικόνων χρησιμοποιήθηκε η βιβλιοθήκη “cv2”. Για ευκολία και οικονομία χρόνου επιλέχθηκε η αποθήκευση των εικόνων μετά την αλλαγή διαστάσης ώστε να μπορούμε να δούμε την επεξεργασμένη εικόνα που τροφοδοτούμε στο δίκτυο. Επιπλέον, χρησιμοποιήθηκε και μια μέθοδος επαύξησης των εικόνων για καλύτερη σύγκλιση των μοντέλων. Συγκεκριμένα χρησιμοποιήθηκε η κλάση “ImageDataGenerator” του “tensorflow” για την δημιουργία επαυξημένων εικόνων με βάση τρεις παράγοντες οι οποίοι είναι: περιστροφή, μετατόπιση πλάτους, μετατόπιση ύψους. Για την καλύτερη επίδοση των μοντέλων στην φάση εκπαίδευσης χρησιμοποιήθηκε η κλάση “ReduceLROnPlateau” η οποία ελέγχει αν ένα μέτρο επίδοσης (στην δική μας περίπτωση το σφάλμα) βελτιώθηκε σε ένα εύρος εποχών και στην περίπτωση που δεν βελτιώθηκε μειώνει των συντελεστή εκπαίδευσης με βάση ένα παράγοντα (π.χ. 0.001). Άλλη μια μέθοδος “κουρδίσματος” των παραμέτρων που χρησιμοποιήθηκε είναι η κλάση “ModelCheckpoint” η οποία μας βοηθά να κρατήσουμε τα βάρη που διαμορφώθηκαν στην φάση εκπαίδευσης σε ένα εύρος εποχών για επαναχρησιμοποίηση στο μέλλον. Τέλος για όλα τα μοντέλα συνάρτηση κόστους επιλέχθηκε η κατηγορηματική σταυροεντροπία.

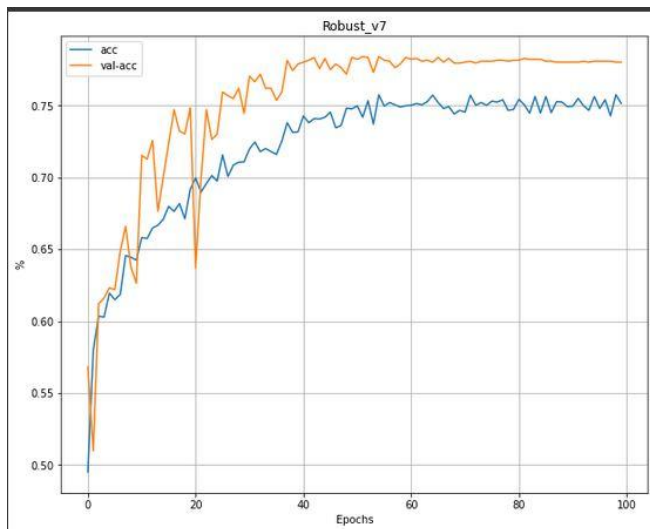
3. ROBUST MODEL

Το πρώτο μοντέλο που κατασκευάστηκε αποτελείται από βαθιά στρώματα συνελιζων καθώς και αρκετά πλήρως συνδεδεμένα στρώματα. Μεταξύ των συνελκτικών στρωμάτων υπάρχουν κάποια “pooling” στρώματα καθώς και κάποια στρώματα κανονικοποίησης. Σύμφωνα με τους κατασκευαστές του το δίκτυο είναι ιδανικό για συναισθηματική ανάλυση εικόνων βέβαια αυτό δεν σημαίνει πως δεν μπορεί να χρησιμοποιηθεί για κλασσική ταξινόμηση εικόνων. Αναλυτικότερα, η εικόνα εισόδου (διάστασης 244,224) επεξεργάζεται από δύο συνελκτικά στρώματα τα οποία στέλνουν την έξοδο τους σε “max-pooling” και “normalization” στρώματα. Το πρώτο

συνελικτικό στρώμα αποτελείται από 96 φίλτρα διάστασης (11,11) και βήμα(stride) 4 pixel. Το δεύτερο συνελικτικό στρώμα αποτελείται από 256 φίλτρα διαστασης (5,5) και βήμα 2 pixel . Έπειτα ακολουθούν τα πλήρως συνδεδεμένα στρώματα. Το σύνολο των παραμέτρων αυτού του μοντέλου είναι 3,639,267.

Σύμφωνα με τα λεγόμενα του κατασκευαστή [1] ένα ενδιαφέρον γεγονός για αυτό το μοντέλο είναι πως το τελευταίο του πλήρως συνδεδεμένο στρώμα πριν το στρώμα εξόδου η διάσταση του τένσορα εξόδου είναι 24 όσα είναι και τα συναισθήματα του ανθρώπου, τα οποία χωρίζονται σε μια λίστα με θετικά και μία με αρνητικά.

Όσον αφορά την διαδικασία εκπαίδευσης σε αυτό το μοντέλο η ακρίβεια και τα σφάλματα στο υποσύνολο εκπαίδευσης κυμαίνονται σε ένα εύρος 75-80% και 0.5 αντίστοιχα ενώ στο υποσύνολο επικύρωσης η ακρίβεια βρίσκεται σε λίγο χαμηλότερα επίπεδα (76-79%) και το σφάλματα (0.52) αντίστοιχα . Επομένως μπορούμε να πούμε πως αυτό το μοντέλο ξεπέρασε το πρόβλημα του over-fitting. Στο σχήμα 1 μπορείτε να δείτε και το γράφημα που απεικονίζει την ακρίβεια εκπαίδευσης και την ακρίβεια επικύρωσης.[1]



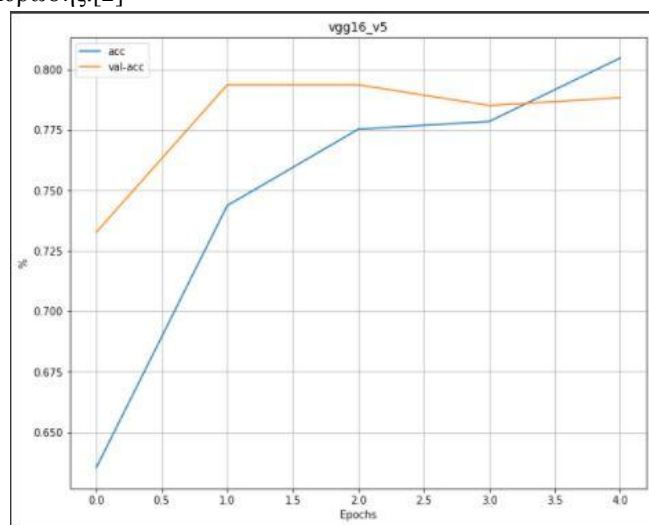
Σχήμα 1: Γράφημα ακρίβειας εκπαίδευσης και ακρίβειας επικύρωσης του μοντέλου Robust

4. VGG-16 MODEL

Ένας ακόμη τρόπος αποδοτικής ταξινόμησης εικόνων είναι η μέθοδος του Mobile Learning. Αυτή η μέθοδος γίνεται δυνατή μέσω του framework “keras” το οποίο μας δίνει την δυνατότητα να χρησιμοποιήσουμε τα προ-εκπαιδευμένα βάρη του γνωστού VGG δικτύου για το δικό μας πρόβλημα ταξινόμησης. Σχετικά με την αρχιτεκτονική του VGG , η εικόνα εισόδου (υποχρεωτικά

διάστασης (224,224)) εισάγεται σε μία στοίβα συνελικτικών στρωμάτων με υποδεκτικό πεδίο (3,3) και βήμα 1 pixel ,με ενδιάμεσα στρώματα “pooling” και “normalization” . Έπειτα ακολουθεί μια στοίβα απο πλήρως συνδεδεμένα στρώματα . Αυτή η ακολουθία στοίβας στρωμάτων συνεχίζεται , σε κάθε επίπεδο με διαφορετικές τιμές φίλτρων ,υποδεκτικών πεδίων και βήματων. Επιπλέον προστέθηκαν 2 ακόμη πλήρως συνδεδεμένα στρώματα και ένα στρώμα “pooling”. Όσον αφορά την διαδικασία εκπαίδευσης σε αυτό το μοντέλο επιλέχθηκε μικρός αριθμός εποχών (διότι κατέχουμε προ-εκπαιδευμένα βάρη) .

Η ακρίβεια που πέτυχε αυτο το μοντέλο στην διαδικασία εκπαίδευσης κυμάνθηκε στο εύρος (75-77%) για το υποσύνολο εκπαίδευσης και (76-79%) για το υποσύνολο επικύρωσης.[2]



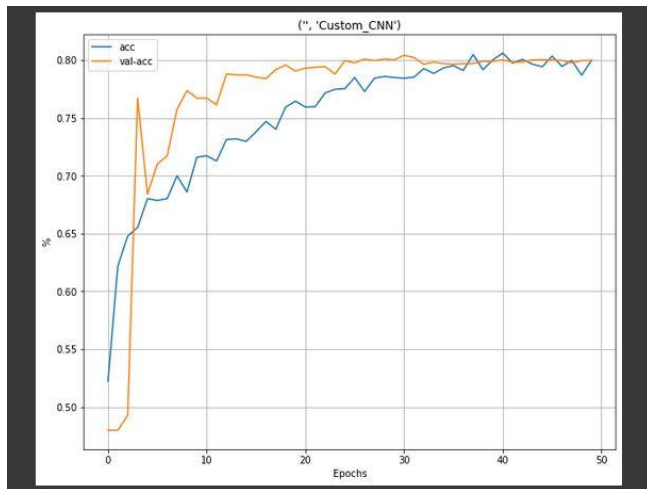
Σχήμα 2: Γράφημα ακρίβειας εκπαίδευσης και ακρίβειας επικύρωσης του μοντέλου VGG

5. CUSTOM CNN MODEL

Αυτό το μοντέλο χρησιμοποιεί την λογική στοίβας που έχει και το VGG . Συγκεκριμένα έχουμε μια “συνελικτική” στοίβα και μία στοίβα πλήρως συνδεδεμένων στρωμάτων. Η “συνελικτική” στοίβα αποτελείται από ένα στρώμα συνέλιξης με διακριτό αριθμό φίλτρων διάστασης (3,3) και στρώματα “pooling” και “normalization” . Σχετικά με την αρχιτεκτονική του δικτύου , η εικόνα εισόδου εισέρχεται σε ένα συνελικτικό στρώμα με 16 φίλτρα διάστασης (3,3) έπειτα ακολουθεί ένα στρώμα “pooling” και έπειτα έχουμε μία “συνελικτική” στοίβα με 32 ,64 ,128 φίλτρα και τέλος έχουμε 3 στοίβες πλήρως συνδεδεμένων στρωμάτων .

Σχετικά με την εκπαίδευση του μοντέλου η μέγιστη ακρίβεια είναι ίδια και στα δύο υποσύνολα εκπαίδευσης 80% καθώς και το ελάχιστο σφάλμα 0.6 . Στο παρακάτω

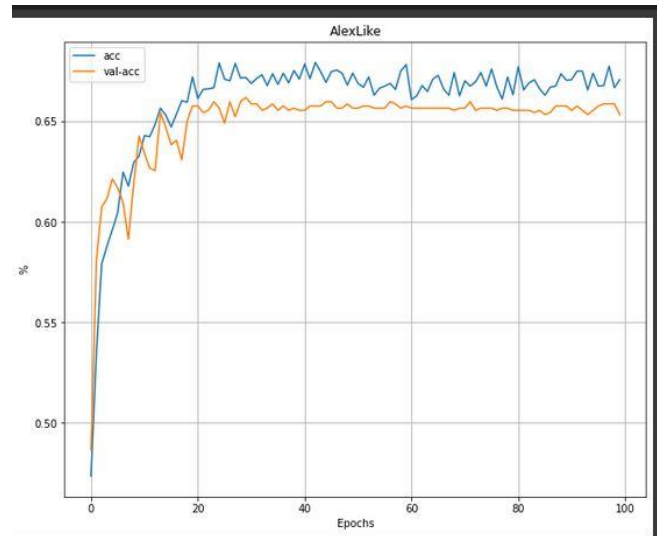
σχήμα μπορείτε να δείτε και την καμπύλη της ακολουθίας της ακρίβειας εκπαίδευσης και επικύρωσης



Σχήμα 3: Γράφημα ακρίβειας εκπαίδευσης και ακρίβειας επικύρωσης του μοντέλου Custom CNN

6. ALEX LIKE MODEL

Το τελευταίο μοντέλο που κατασκευάστηκε έχει παρόμοια αρχιτεκτονική με το γνωστό σε όλους “AlexNet” το οποίο ήταν το πρώτο μοντέλο που στον γνωστό διαγωνισμό “ImageNet” ξεπέρασε κατά 10% το “top-5” σφάλμα του προηγούμενου νικητή του διαγωνισμού. Οι παραλλαγές που έχουν γίνει αφορούν τις διαστάσεις των εικόνων εισόδου που στην περίπτωση μας είναι (224,224) ενώ στο επίσημο “AlexNet” είναι (227,227). Επιπλέον, μια ακόμη παραλλαγή που έγινε είναι η πρόσθεση ενός ακόμη συνελκτικού στρώματος με αριθμό φίλτρων 16 διάστασης (11,11) πριν το επίσημο πρώτο συνελκτικό στρώμα του AlexNet το οποίο έχει 96 φίλτρα διάστασης (1,1). Έπειτα έχουμε 4 μπλοκ σύνελκτικών στρωμάτων με 192 φίλτρα διάστασης (1,1) με ενδιάμεσο “pooling” και ένα συνελκτικό μπλοκ με 10 φίλτρα διάστασης (1,1). Τέλος, έχουμε τρία πλήρως συνδεδεμένα στρώματα. Η βασική του διαφορά με το επίσημο AlexNet είναι ο αριθμός των φίλτρων στα συνελκτικά μπλοκ αλλά η διάσταση τους είναι ίδια.



Σχετικά με την εκπαίδευση, το μοντέλο πέτυχε παρόμοια ποσοστά με τα προηγούμενα και συγκεκριμένα η ακρίβεια εκπαίδευσης κυμαίνεται 74% ενώ η ακρίβεια επικύρωσης κυμαίνεται γύρω στο 76% όπως φαίνεται και στο σχήμα 4 και αντίστοιχα τα σφάλματα στο 0.58 και 0.57.

5.1 ΣΥΜΠΕΡΑΣΜΑΤΑ

Ένα κοινό χαρακτηριστικό που παρατηρήθηκε είναι πως σε όλα τα μοντέλα (εκτός από αυτό που χρησιμοποιήθηκε Mobile Learning) είναι πως η ακρίβεια και το σφάλμα επικύρωσης μετά από κάποιο αριθμό εποχών, διαφορετικό σε κάθε μοντέλο, “φράζουν” δηλαδή η τιμή τους ταλαντώνεται γύρω από έναν αριθμό (π.χ. στο 77% η ακρίβεια του Robust μοντέλου). Αυτό το χαρακτηριστικό παρατηρήθηκε και σε εκδοχές εκπαίδευσης με 50 και 100 εποχές καθώς και με διαφορετικές κατανομές συντελεστή εκπαίδευσης. Βέβαια η εκδοχή που πέτυχαν καλύτερη ακρίβεια στην φάση ελέγχου ήταν με 100 εποχές και συντελεστή βαρύτητας 10^{-3} με “step decay” 10^{-6} .

7. ΑΠΟΤΕΛΕΣΜΑΤΑ

Τα καλύτερα αποτελέσματα του Robust Μοντέλου έγιναν στην εκδοχή εκπαίδευσης με 100 εποχές και αρχικό συντελεστή εκπαίδευσης 10^{-3} . Αυτό το μοντέλο πέτυχε και την μεγαλύτερη ακρίβεια από όλα τα υπόλοιπα (~80%).

Το καλύτερο ποσοστό ακρίβειας που πέτυχε το μοντέλο Custom CNN ήταν ~78% σε ίδιες τιμές εποχών και συντελεστή βαρύτητας. Επίσης είχε παρόμοια απόδοση με την δεύτερη καλύτερη εκδοχή του μοντέλου Robust. Σχετικά με τα μοντέλα VGG και AlexLike μπορούμε να πούμε πως μέθαν τα δεδομένα με παρόμοιο τρόπο καθώς πετύχαν πολύ κοντινές αποδόσεις ακρίβειας (και στα δύο μοντέλα η ακρίβεια ήταν γύρω στο 77%). Η κατάταξη της ακρίβειας

τον μοντέλο στη φάση ελέγχου σε όλες τις διαφορετικές εκδοχές των μοντέλων ήταν :

1. Robust ~80%
2. Custom CNN ~79%
3. AlexLike 77,7%
4. VGG 77,4%

8. REFERENCES

[1] You, Quanzeng, et al. "Robust image sentiment analysis using progressively trained and domain transferred deep networks." *Twenty-ninth AAAI conference on artificial intelligence*. 2015.

[2] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks (AlexNet) ImageNet Classification with Deep Convolutional Neural Networks (AlexNet) ImageNet Classification with Deep Convolutional Neural Networks." *ImageNet Classification with Deep Convolutional Neural Networks*.