

Σχήμα 19.41: Απόσπασμα από οπτικοποίηση ενός CNN ταξινόμησης χειρόγραφων ψηφίων (Adam W. Harley). Διακρίνεται η συνέλιξη δεδομένων από 4 pooling επίπεδα του προηγούμενου σταδίου, στο 2<sup>ο</sup> επίπεδο συνέλιξης.

Τέλος, σε επίπεδα συνέλιξης πέρα από αυτό της εισόδου του CNN, συνηθίζεται να γίνεται συνέλιξη μεταξύ των επιμέρους επιπέδων που προκύπτουν σε προηγούμενα στάδια. Με αυτό τον τρόπο το δίκτυο, σε βαθύτερα επίπεδα, συνδυάζει χαρακτηριστικά λεπτότερης υφής των προηγούμενων επιπέδων και συνθέτει χαρακτηριστικά περισσότερο αδρής υφής (Σχήμα 19.41).

### Στάδιο σμίκρυνσης

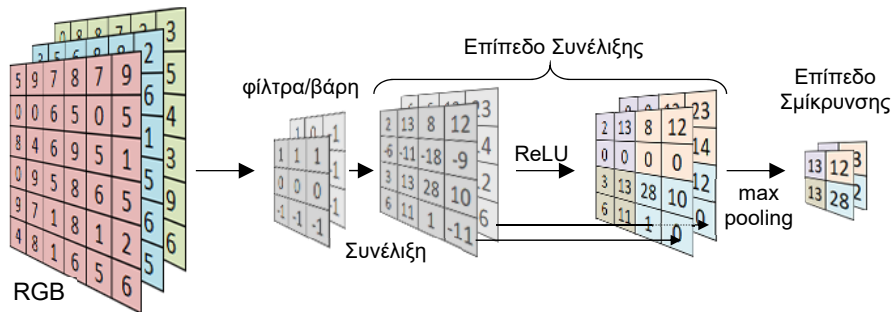
Στα CNN συνηθίζεται μεταξύ δύο ή μετά από μερικά διαδοχικά επίπεδα συνέλιξης, να παρεμβάλλεται ένα επίπεδο σμίκρυνσης (ή συγκέντρωσης/pooling layer). Τα είδη των υπολογισμών που συνηθίζονται στα επίπεδα σμίκρυνσης είναι το μέγιστο (max pooling) και ο μέσος όρος (average pooling). Στο Σχήμα 19.42 απεικονίζονται 2 παραδείγματα τέτοιων υπολογισμών σε πίνακα 4×4 με παράθυρο pooling 2×2 και διασκελισμό 2. Στο παράδειγμα αριστερά (max pooling) είναι  $28 = \max(28, 10, 1, 0)$  ενώ στο παράδειγμα δεξιά (average pooling) είναι  $7.5 = (2+13+0+0)/4$ .



Σχήμα 19.42: Υπολογισμοί max και average pooling σε πίνακα 4×4 με παράθυρο pooling 2×2 και διασκελισμό 2.

Δεδομένου ότι σε ένα στάδιο συνέλιξης εφαρμόζονται συνήθως πολλά φίλτρα και παράγονται πολλά επίπεδα, το στάδιο σμίκρυνσης που τυχόν ακολουθεί εφαρμόζεται σε όλα αυτά τα επίπεδα. Στο Σχήμα 19.43 η συνέλιξη παράγει μια ομάδα από 2 επίπεδα (επειδή υπάρχουν 2 φίλτρα) στις τιμές των οποίων προστίθενται τιμές πόλωσης και η συνολική τιμή περνά μέσα από τη συνάρτηση ενεργοποίησης (ReLU) του νευρώνα. Στη συνέχεια υπάρχει ένα επίπεδο σμίκρυνσης που μετατρέπει τον 4×4 πίνακα σε 2×2. Το επίπεδο σμίκρυνσης σε αυτό το παράδειγμα παράγεται μέσω ενός παραθύρου διαστάσεων 2×2 που διατρέχει οριζόντια την έξοδο του επιπέδου συνέλιξης με διασκελισμό (stride) 2. Η τιμή που προκύπτει για κάθε εφαρμογή του παραθύρου συγκέντρω-

σης στο Σχήμα 19.43 ισούται με το μέγιστο από τις εκάστοτε 4 τιμές, καθώς πρόκειται για επίπεδο σμίκρυνσης μεγίστου (max pooling layer).



Σχήμα 19.43: Εικόνα RGB ακολουθούμενη από επίπεδο συνέλιξης και επίπεδο σμίκρυνσης, στην αρχή ενός CNN.

Μέσω των επιπέδων σμίκρυνσης επιτυγχάνονται τα ακόλουθα:

- Όπως και στη σμίκρυνση εικόνων, μόνο τα κύρια (έντονα) χαρακτηριστικά των επιπέδων διατηρούνται. Απλοποιείται έτσι η αναπαράσταση.
- Μειώνεται η διάσταση των επιπέδων, άρα και το πλήθος των παραμέτρων που πρέπει να υπολογιστούν, συμβάλλοντας έτσι στην αποφυγή υπερπροσαρμογής αλλά και σε ταχύτερο υπολογισμό.
- Το δίκτυο είναι πιο ανθεκτικό σε μικρές παραμορφώσεις/αλλοιώσεις στην είσοδο καθώς αυτές δεν επηρεάζουν ιδιαίτερα το μέγιστο ή το μέσο όρο μιας μικρής περιοχής.
- Επιτυγχάνεται μια αναπαράσταση που επιτρέπει τον εντοπισμό αντικειμένων ασχέτως του που βρίσκονται στην εικόνα.

### Πλήρως διασυνδεδεμένα επίπεδα

Στο τελικό στάδιο επεξεργασίας ενός CNN υπάρχουν πάντα πλήρως διασυνδεδεμένα επίπεδα που υλοποιούν έναν ταξινομητή. Συνήθως υπάρχουν 2 τέτοια επίπεδα και ένα τελικό τύπου *softmax*. Δουλειά του ταξινομητή είναι να αποφανθεί για την κλάση του αντικειμένου (ή αντικειμένων) που απεικονίζονται στην εικόνα με βάση τα διάφορα χαρακτηριστικά που εντοπίστηκαν στα προγενέστερα στάδια επεξεργασίας. Προφανώς τα επίπεδα αυτά είναι μιας διάστασης και επομένως θα πρέπει οι πίνακες δύο διαστάσεων του προηγούμενου σταδίου επεξεργασίας να μετατραπούν σε πίνακα μιας διάστασης. Ο πιο απλός τρόπος είναι αυτός που απεικονίζεται στο Σχήμα 19.14. Τυχόν περισσότερα επίπεδα αυξάνουν το πλήθος των στοιχείων στη μία διάσταση. Άλλος τρόπος είναι να γίνει ένα τελευταίο pooling με ταυτόχρονη μετάβαση σε μία διάσταση.

### Πρακτικά ζητήματα στη χρήση ConvNets

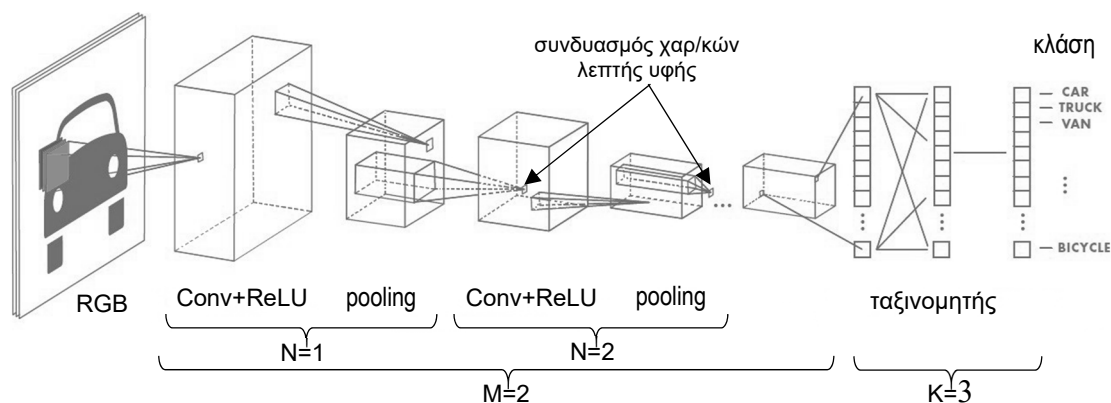
#### Γενική δομή CNN

Η γενική δομή ενός CNN είναι η ακόλουθη:

INPUT -> [[CONV -> RELU]\*N -> POOL?]\*M -> [FC -> RELU]\*K -> FC

Δηλαδή μετά την είσοδο, ακολουθούν N διαδοχικά στάδια συνέλιξης με νευρώνες ReLU, προαιρετικά ακολουθούμενα από στάδιο pooling. Αυτός ο συνδυασμός, με ή

χωρίς pooling επίπεδο, μπορεί να επαναλαμβάνεται  $M$  φορές. Στο τέλος υπάρχουν  $K$  πλήρως διασυνδεδεμένα επίπεδα που καταλήγουν (με πλήρη διασύνδεση) σε *softmax* επίπεδο. Τυπικές τιμές για τα  $N$ ,  $M$  και  $K$  είναι μικροί ακέραιοι, συνήθως ως 3, χωρίς να αποκλείονται και πολύ μεγαλύτερες τιμές, ανάλογα με το πρόβλημα. Παράδειγμα ολοκληρωμένου CNN απεικονίζεται στο Σχήμα 19.44. Τονίζεται ξανά ο συνδυασμός χαρακτηριστικών από περισσότερα του ενός πρωθύστερα pooling επίπεδα, στα διάδοχα επίπεδα συνέλιξης. Συνήθως εμπλέκονται όλα τα επίπεδα και αφήνεται στην εκπαίδευση να καθορίσει τον ακριβή τρόπο.



Σχήμα 19.44: Σχηματική αναπαράσταση ενός CNN με βάση τη γενική δομή.

Για να δοθεί μια ρεαλιστική εικόνα βαθέως συνελκτικού δικτύου, δίνονται στη συνέχεια μερικές λεπτομέρειες από το *AlexNet* του Alex Krizhevsky, το πρώτο ConvNet που νίκησε στον ετήσιο διαγωνισμό αναγνώρισης εικόνων IVL SRC, εκείνον του 2012. Η νίκη αυτή ήταν αξιομνημόνευτη καθώς βελτίωσε αλματωδώς (κατά τουλάχιστον 10 ποσοστιαίες μονάδες) το προηγούμενο καλύτερο σκορ και κατέστησε "εν μία νυκτί" παρωχημένες όλες τις μη βασισμένες σε συνελκτικά δίκτυα προσεγγίσεις.

Το ALEXNET περιείχε 5 συνελκτικά και 3 max pooling στάδια επεξεργασίας, ενώ ο ταξινομητής του αποτελούταν από 2 πλήρως συνδεδεμένα στρώματα και ένα τελικό *softmax* με 1000 νευρώνες, καθώς το πρόβλημα ήταν ταξινόμηση σε 1000 κλάσεις. Με είσοδο RGB εικόνες διαστάσεων  $224 \times 224$ , το ALEXNET περιείχε περίπου 670 χιλιάδες νευρώνες και 60 εκατομμύρια προσαρμοζόμενα βάρη!

### Μεταφορά εκπαίδευσης/μάθησης

Τα CNN επιτυγχάνουν πολύ καλή επίδοση στην αναγνώριση αντικειμένων από εικόνες. Βέβαια, είναι γενικά δύσκολα και αργά στην εκπαίδευσή τους και απαιτούν πολλά δεδομένα. Για τον λόγο αυτό συχνά χρησιμοποιούνται προεκπαιδευμένα μοντέλα (π.χ. RESNET50) σε μεγάλες βάσεις δεδομένων (π.χ. IMAGENET). Σε αυτά, αντικαθίστανται πιθανώς τα τελικά επίπεδα που συνθέτουν τον ταξινομητή και επανεκπαιδεύονται στα δεδομένα του προβλήματος, παγώνοντας τα βάρη στα εμπρός επίπεδα. Εφόσον τα δεδομένα του προβλήματος δεν είναι πολύ διαφορετικά από αυτά που εκπαίδευσαν τον αρχικό CNN, αυτή η τεχνική της μεταφοράς της μάθησης μεταξύ ΤΝΔ (*transfer learning*) εξασφαλίζει πολύ καλές επιδόσεις σε σύντομο χρονικό διάστημα. Αν τα δεδομένα είναι άλλης φύσεως, το CNN πρέπει να εκπαιδευτεί εξ αρχής σε όλα του τα επίπεδα.