0.1 Σχετική βιβλιογραφία

Σε αυτή την ενότητα, παρουσιάζουμε ορισμένες από τις πιο συναφείς μεθόδους για την εργασία μας και άλλες που μελετήθηκαν για τον σχεδιασμό αυτής της προσέγγισης. Οι μέθοδοι αυτές χωρίζονται σε δύο ενότητες *Αναγνώριση Δραστηριότητας* και *Εντοπισμός Δραστηριότητας*. Το πρώτο μέρος αναφέρεται σε κλασικές μεθόδους ταξινόμησης δράσης που εισήχθησαν μέχρι πρόσφατα και το δεύτερο μέρος, αντίστοιχα, σε πρόσφατες μεθόδους εντοπισμού της δράσης.

0.1.1 Αναγνώριση Δραστηριότητας

Οι πρώτες προσεγγίσεις για την κατάταξη της δράσης αποτελούνταν από δύο βήματα α) αρχικά υπολογισμός σύνθετων 'χειροποίητων' χαρακτηριστικών από ακατέργαστα καρέ βίντεο και β) εκπαίδευση ενός ταξινομητή με βάση αυτά τα χαρακτηριστικά. Αυτά τα χαρακτηριστικά μπορούν να διαχωριστούν σε 3 κατηγορίες: 1) προσεγγίσεις χωροχρονικού όγκου (space-time volume), 2) τροχιές (trajectories) και 3) χωροχρονικά χαρακτηριστικά. Για τις μεθόδους χωροχρονικού όγκου, η προσέγγιση είναι η εξής: Με βάση τα training βίντεο, το σύστημα συνάπτει ένα μοντέλο τρισδιάστατου χωροχρόνου, συνενώνοντας δυσδιάστατες εικόνες (διάσταση x-y) κατά τη διάρκεια του χρόνου (διάσταση t ή z), για την αναπαράσταση κάθε δράσης. Όταν το σύστημα δέχεται ένα βίντεο που δεν έχει ετικέτα, κατασκευάζει μια τρισδιάστατη χωροχρονική αναπαράσταση που αντιστοιχεί σε αυτό το βίντεο. Αυτό η νέα τρισδιάστατη αναπαράσταση, στη συνέχεια, συγκρίνεται με κάθε μοντέλο 3D χωροχρόνου, συγκρίνοντας την ομοιότητα στο σχήμα και την εμφάνιση μεταξύ αυτών των δύο χωροχρονικών όγκων. Το σύστημα εξάγει την κατηγορία του άγνωστου βίντεο, αντιστοιχώντας την με αυτήν της δράσης με την υψηλότερη ομοιότητα. Επιπλέον, υπάρχουν διάφορες παραλλαγές των χωροχρονικών αναπαραστάσεων. Αντί της αναπαράστασης space-time volume, το σύστημα μπορεί να αναπαριστά τη κάθε δράση ως τροχιές σε χωροχρονικές διαστάσεις ή ακόμη περισσότερο, η ενέργεια μπορεί να αναπαρασταθεί ως ένα σύνολο χαρακτηριστικών που εξάγονται από τον χωροχρονικό όγκο ή τις τροχιές. Οι 'καθαρές' χωροχρονικές αναπαραστάσεις περιλαμβάνουν μεθόδους σύγκρισης των περιοχών προσκηνίου ενός ατόμου (δηλ. σιλουέτες) όπως Bobick and Davis 2001,συγκρίνοντας όγκους σε σχέση με επιφάνεια τους όπως οι Shechtman and Irani 2005. Η μέθοδος Ke, Sukthankar, and Hebert 2007χρησιμοποιεί oversegmented όγκους, αυτομάτως υπολογίζοντας ένα σύνολο τμημάτων τρισδιάστατου όγκου XYT που αντιστοιχεί σε έναν κινούμενο άνθρωπο. Rodriguez, Ahmed, and Shah 2008 πρότειναν φίλτρα για να αποτυπόνουν τα χαρακτηριστικά του χωροχρονικού όγκου, προκειμένου να τα ταιριάζουν πιο αξιόπιστα και αποδοτικά. Από την άλλη πλευρά, οι προσεγγίσεις με βάση την τροχιά περιλαμβάνουν την αναπαράσταση μιας ενέργειας ως σύνολο 13 κοινών διαδρομών (Sheikh, Sheikh, and Shah 2005) ή τη χρήση ενός συνόλου $XYZT$-διαστάσεων κοινών τροχιών που λαμβάνονται από κινούμενες κάμερες (Yilmaz and Shah 2005). Τέλος, διάφορες μέθοδοι χρησιμοποιούν τοπικά χαρακτηριστικά που εξάγονται από χωροχρονικούς όγκους τριών διαστάσεων, όπως η εξαγωγή τοπικών χαρακτηριστικών σε κάθε καρέ του βίντεο και η ένωση του χρονικά (Chomat and Crowley 1999; Zelnik-Manor and Irani 2001; Blank

et al. 2005, η εξαγωγή αραιών χωροχρονικών τοπικών σημείων ενδιαφέροντος από τρισδιάστατους όγκους (Laptev and Lindeberg 2003; Dollar et al. 2005; Niebles, Wang, and Li 2006; Alper Yilmaz and Mubarak Shah 2005; Ryoo and Aggarwal 2006) Οι προσεγγίσεις αυτές κατέστησαν την επιλογή των χαρακτηριστικών σημαντικό παράγοντα για την απόδοση του δικτύου. Αυτό συμβαίνει επειδή οι διαφορετικές κατηγορίες δράσεν μπορεί να διαφέρουν δραματικά από την άποψη της εμφάνισής τους και των μοτίβων κίνησης. Ένα άλλο πρόβλημα ήταν ότι οι περισσότερες από αυτές τις προσεγγίσεις κάνουν υποθέσεις, υπό τις οποίες το βίντεο λήφθηκε λόγω προβλημάτων όπως το γεμάτο φόντο, γωνιές κάμερας κλπ. Μια ανασκόπηση των τεχνικών, που χρησιμοποιούνταν μέχρι το 2011, παρουσιάζεται στο Aggarwal and Ryoo 2011.

Τα πρόσφατα αποτελέσματα σε βαθιές αρχιτεκτονικές και ειδικά στον τομέα της ταξινόμηση εικόνας έδωσε κίνητρο στους ερευνητές να εκπαιδεύσουν δίκτυα CNN για το πρόβλημα της αναγνώρισης δράσης. Η πρώτη σημαντική απόπειρα έγινε από τους Karpathy et al. 2014. Σχεδίασαν την αρχιτεκτονική τους με βάση το καλύτερο CNN στον διαγωνισμό ImageNet. Εξερευνούν διάφορες μεθόδους για τη σύντηξη των χωροχρονικών λειτουργιών χρησιμοποιώντας δυσδιάσατες διαδικασίες κυρίως και τρισδιάστατη συνέλιξη μόνο σε αργή σύντηξη. Οι Simonyan and Zisserman 2014 χρησιμοποιήσαν 2 ΝNς, ένα για χωρικές πληροφορίες και ένα για οπτική ροή και τα συνδύασαν με τη χρήση της καθυστερημένης σύντηξης. Δείχνουν ότι η εξόρυξη χωρικού περιεχομένου από τα βίντεο και περιεχόμενο κίνησης από την οπτική ροή μπορεί να βελτιώσει σημαντικά την ακρίβεια της αναγνώρισης της δράσης. Οι Feichtenhofer, Pinz, and Zisserman 2016 επέκτειναν αυτή την προσέγγιση με τη χρήση πρώιμης σύντηξης στο τέλος των convolutional layeers αντί για καθυστερημένης σύντηξης, η οποία λαμβάνει χώρα στο τελευταίο επίπεδο του δικτύου. Πάνω σ΄ αυτό, χρησιμοποίησαν ένα δεύτερο δίκτυο για το χρονικό περιεχόμενοοοο το οποίο συνδέουν με το το άλλο δίκτυο με χρήση της καθυστερημένης σύντηξης. Επιπλέον, οι Wang et al. 2016 στήριξαν την μέθοδος τους σε αυτήν που πρότειναν οι Simonyan and Zisserman 2014. Ασχολούνται με το πρόβλημα του την εύρεσης χρονικού περιεχομένου και εκπαιδεύουν το δίκτυο τους, παρέχοντας του λίγα δείγματα. Η προσέγγισή τους, την οποία ονόμασαν Temporal Segment Network (TSN), διαχωρίζει το βίντεο εισόδου σε Κ τμήματα και ένα σύντομο απόσπασμα από κάθε τμήμα επιλέγεται για ανάλυση. Στη συνέχεια, συνδέουν το εξαγόμενο χωροχρονικό περιεχόμενο, πραγματοποιώντας τελικά την πρόβλεψή τους. Πιο πρόσφατα, οι Zhang et al. 2016 και οι Zhu et al. 2017 χρησιμοποίησαν την two-stream , επίσης. Οι Zhang et al. 2016 αντικατέστησαν την οπτική ροή με ένα διάνυσμα κίνησης που μπορεί να ληφθεί απευθείας από τα συμπιεσμένα βίντεο χωρίς επιπλέον υπολογισμό και το τροφοδοτούν στο δίκτυο. Οι Zhu et al. 2017 εκπαίδευσαν ένα CNN για τον υπολογισμό της οπτικής ροής, καλώντας το MotionNet και χρησιμοποίησαν ένα CNN ως χρονικό stream για προβάλλουν τις πληροφορίες κίνησης έργου σε κατηγορίες δράσεωνο. Τέλος χρησιμοποιούν Την καθυστερημένη σύντηξη μέσω της μέσης τιμής με βάρη των σκορ πρόβλεψης των χρονικών και χωρικών stream. Από την άλλη πλευρά, μια νέα προσέγγιση εισήχθη από τους Γιρδηαρ και Ραμαναν 2017ενσωματώνοντας χάρτες προσοχής με σκοπόν να βελτιώσουν σημαντικά την απόδοση της αναγνώρισης δράσης.

Ορισμένες άλλες μέθοδοι περιλάμβαναν ένα δίκτυο RNN ή LSTM για την ταξινόμηση κάνουν οι Donahue et al. 2017, οι Joe Yue-Hei Ng et al. 2015 και οι Ma et al. 2017. Οι Donahue et al. 2017αντιμετωπίζουν τη πρόκλησης των μεταβλητών μήκη των ακολουθιών εισόδου και εξόδου, εκμεταλλευόμενοι τα convolutional layers και μεγάλου εύρους χρονικές αναδρομές (recursions). Προτείνουν ένα Long-term Recurrent Convolutional Network (LRCN), το οποίο είναι ικανό να αντιμετωπίσει τις εργασίες αναγνώρισης, λεζάντας εικόνας και περιγραφής βίντεο. Για να ταξινομήσουν μια δεδομένη ακολουθία καρέ, το LRCN λαμβάνει αρχικά ως είσοδο ένα καρέ, και πιο συγκεκριμένα τα κανάλια RGB και την οπτική ροή του, και προβλέπει μια ετικέτα. Μετά από αυτό, εξάγει την κλάση του βίντεο μέσω του μέσου όρου των πιθανοτήτων των ετικετών, επιλέγοντας την πιο πιθανή κλάση. Οι Joe Yue-Hei Ng et al. 2015 πρώτα διερευνούν διάφορες προσεγγίσεις για χρονική ομαδοποίησης (temporal pooling) των χαρακτηριστικών. Αυτές οι τεχνικές περιλαμβάνουν τον χειρισμό καρέ βίντεο ξεχωριστά από 2 αρχιτεκτονικές CNN: είτε απ΄ το AlexNet είτε απ΄ το GoogleNet, και αποτελούνται από πρώιμη σύντηξη, καθυστερημένη σύντηξη και ενός συνδυασμού αυτών. Επιπλέον, προτείνουν ένα RNN προκειμένου να εξετάσουν τα βίντεο κλιπ ως ακολουθίες ενεργοποιήσεων ̃NN. Το προτεινόμενο LSTM λαμβάνει ως είσοδο την έξοδο του τελικού CNN layer για κάθε συνεχόμενο καρέ και μετά από 5 LSTM layers και χρησιμοποιώντας έναν softmax ταξινομητή, προτείνει μία ετικέτα. Για την ταξινόμηση του βίντεο, επιστρέφουν μια ετικέτα μετά το τελευταίο βήμα, εφαρμόζουν max-pooling στις προβλέψεις στην διάσταση του χρόνου, αθροίζουν τις προβλέψεις στην διάσταση του χρόνου και επιστρέφουν το μέγιστο ή έναν γραμμικό συνδυασμό με βάρη των προβλέψεων υπό έναν παράγοντα g, τα αθροίζουν και επιστρέφουν το μέγιστο. Έδειξαν ότι όλες οι προσεγγίσεις είναι 1% διαφορετικές με προκατάληψη για τη χρήση των προβλέψεων με βάρη για την υποστήριξη της ιδέας ότι το LSTM γίνεται προοδευτικά πιο ενημερωμένο. Τελευταίοι αλλά όχι λιγότερο σημαντικοι, οι Ma et al. 2017 χρησιμοποίησαν ένα two-stream ConvNet για εξαγωγή χαρακτηριστικών και είτε ένα LSTM ή convolutional πάνω από το χρονικώς κατασκευασμένους πίνακες χαρακτηριστικών για τη σύντηξη χωρικών και χρονικών πληροφοριών. Χρησιμοποιούν ένα ResNet-101 για την εξόρυξη χαρτών ενεργοποίησης τόσο για χωρικές όσο και για χρονικές ροές. Χωρίζουν το βίντεο σε διάφορα τμήματα, όπως έκαναν οι Wang et al. 2016, και χρησιμοποίησαν ένα επίπεδο temporal pooling για την εξαγωγή διακεκριμένων χαρακτηριστικών. Αφού λάβουν αυτά τα χαρακτηριστικά, το LSTM εξάγει ενσωματωμένες δυνατότητες από όλα τα τμήματα.

Επιπλέον, οι Tran et al. 2015 διευρένησαν τα 3D Convolutional δίκτυα (Ji et al. 2013) και εισήγαγαν το C3D δίκτυο που έχει 3D convolutional layers με πυρήνες $3 \times 3 \times 3$. Αυτό το δίκτυο είναι σε θέση να μοντελοποιήσει την εμφάνιση και την κίνηση ταυτόχρονα χρησιμοποιώντας τρισδιάστατες συνελίξεις και μπορεί να χρησιμοποιηθεί ως εξαγωγέας χαρακτηριστικών. Συνδυάζοντας την αρχιτεκτονική δύο ροών και τις τρισδιάστατες συνελίξεις οι ἂρρειρα και Ζισσερμαν 2017 πρότειναν το δίκτυο I3D. Πάνω σ΄ αυτό, οι δημιουργοί τονίζουν τα πλεονεκτήματα της μεταφοράς μάθησης για την εργασία της αναγνώρισης επαναλαμβάνοντας τα δυσδιάστατα προ-εκπαιδευμένα βάρη στην 3η διάσταση. Οι Hara, Kataoka, and Satoh 2017 πρότειναν ένα δίκτυο 3D ResNet για την αναγνώριση δράσης με βάση

τα Residual δίκτυα (ResNet)(He et al. 2016) και διερευνούν την απόδοση των δικτύων ResNet με 3D Convolutional πυρήνες. Από την άλλη, οι Diba et al. 2017 βάσισαν την προσέγγισή τους στα DenseNets (Huang et al. 2017) και επέκτειναν την αρχιτεκτονική του DenseNet χρησιμοποιώντας τρισδιάστατα φίλτρα και pooling πυρήνες αντί για δισδιάστατους, ονομάζοντας αυτή την προσέγγιση ως DenseNet3D. Επιπλέον, εισάγουν το Layer χρονικής μετάβασης (ΤΤΛ), το οποίο συνενώνει χρονικά χάρτες χαρακτηριστικών που εξάγονται σε διαφορετικά χρονικά βάθη και αντικαθιστά το επίπεδο μετάβασης του DenseNet. Παράλληλα, οι Diba et al. 2018 εισήγαγαν ένα νέο χρονικό layer το οποίο μοντελοποιεί μεταβλητούς χρονικούς πυρήνες συνέλιξης. Τελευταίοι αλλά εξίσου σημαντικοί, οι Tran et al. 2018πειραματίστηκαν με διάφορες υπόλοιπες αρχιτεκτονικές Residual δικτύου χρησιμοποιώντας συνδυασμούς 2D και 3Δ ςονολυτιοναλ Λαψερ. Σκοπός τους είναι να δείξουν ότι η 2D χωρική συνέλιξη ακολουθούμενη από 1D χρονική συνέλιξη επιτυγχάνει state-of-the-art αποτελέσματα, ονομάζοντας αυτού του τύπου το layer ως R(2 + 1)D. Πρόσφατα οι Guo et al. 2018 πρότειναν ένα framework που μπορεί να μάθει να αναγνωρίζει μια προηγουμένως αθέατη 3D κλάση δράσης με λίγα μόνο παραδείγματα εκμεταλλευόμενο την εγγενή δομή των 3D δεδομένων μέσω μιας γραφικής αναπαράστασης. Ακόμα πιο λεπτομερή παρουσίαση των τεχνικών αναγνώρισης δράσης που χρησιμοποιήθηκαν μέχρι το 2018 πραγματοποιήθηκε από τους Kong and Fu 2018.

0.1.2 Εντοπισμός Δραστηριότητας

Όπως προαναφέρθηκε, ο εντοπισμός δράσης μπορεί να θεωρηθεί ως προέκταση του προβλήματος εντοπισμού αντικειμένων. Αντί να εξάγουμε δισδιάστατα πλαίσια οριοθέτησης σε μία μόνο εικόνας, ο στόχος των συστημάτων εντοπισμού δράσης είναι να εξάγουν action tubes, τα οποία είναι ακολουθίες πλαισίων οριοθέτησης που περιέχουν μια ενέργεια που εκτελέστηκε. Έτσι, υπάρχουν διάφορες προσεγγίσεις, συμπεριλαμβανομένου συνήθως ενός δικτύου ανιχνευτή αντικειμένων και ενός ταξινομητή.

Οι πρώτες προσεγγίσεις ανίχνευσης αντικειμένων περιλάμβαναν την επέκταση ενός αλγορίθμου πρότασης αντικειμένων σε 3 διαστάσεις. Οι Tian, Sukthankar, and Shah 2013 επέκτείναν τα παραμορφώσιμα (deformable) μοντέλα (Felzenszwalb et al. 2010) με το να αντιμετωπίζουν τις δράσεις ως ξωροχρονικά μοτίβα και δημιούργησαν ένα παραμορφώσιμου μοντέλο για κάθε δράση. Οι Jain et al. 2014 εισήγαγαν την έννοια των tubelets, γνωστά και ως ακολουθίες πλαισίων οριοθέτησης και βάσισαν τη μέθοδό τους σε επιλεκτικό αλγόριθμο αναζήτησης (Uijlings et al. 2013), επεκτείνοντας τα superpixels σε supervoxels για την παραγωγή χωροχρονικών σχημάτων. Απ΄ την άλλη, οι Oneata et al. 2014 επέκτειναν μια τυχαιοποιημένη διαδικασία συγχώνευσης superpixels που χρησιμοποιούταν για που χρησιμοποιούνταν για προτάσεις αντικειμένων, όπως παρουσιάστηκαν απ΄ τους Manen, Guillaumin, and Gool 2013. Οι Yu and Yuan 2015 πρώτα προτείνουν πλαίσια οριοθέτησης για κάθε καρέ με χρήση ενός ανιχνευτή ανθρώπου και κίνησης, ενώ, στη συνέχεια, με τη επιλογή των καλύτερων σε σκορ κουτιών, πρότειναν έναν άπλησ το συνδετικό αλγόριθμο με τη διατύπωση την εργασίας σύνδεσης ως πρόβλημα μέγιστης κάλυψης. Οι Gemert et al. 2015 παράγουν χωροχρονικές προ-

τάσεις κατευθείαν από τις πυκνές τροχειές, οι οποίες επίσης χρησιμοποιήθηκαν για ταξινόμηση. Οι Chen and Corso 2015 δημιουργούν ένα γράφημα χωροχρονικής τροχιάς και επιλέγουν προτάσεις δράσεων που βασίζονται μόνο στην εσκεμμένη κίνηση που εξάγεται από το γράφημα. Οι Soomro, Idrees, and Shah 2015 διαχωρίζουν τα τμήματα βίντεο σε supervoxels και χρησιμοποιούν το περιεχόμενο τους ως χωρική σχέση μεταξύ των supervoxels σε σχέση με την δράση του προσκηνίου. Δημιουργούν ένα γράφημα για κάθε βίντεο, όπου τα συπεροξελς σχηματίζουν τους κόμβους και οι κατευθυνμένες άκρες απεικονίζουν τις χωρικές σχέσεις μεταξύ τους. Κατά τη διάρκεια των δοκιμών, κάνουν μια βόλτα στο περιβάλλον, όπου κάθε βήμα καθοδηγείται από τις σχέσεις περιβάλλοντος κατά τη διάρκεια της εκπαίδευσης, με αποτέλεσμα μια κατανομή πιθανότητας μιας δράσης για όλα τα συπεροξελς. Οι Mettes, Gemert, and Snoek 2016 αντί για τοποθέτηση πλαισίων σε όλα τα καρέ των βίντεο, σχολίασαν σημεία σε ένα αραιό υποσύνολο καρέ του βίντεο και χρησιμοποίησαν προτάσεις που λαμβάνονται μέσω ενός μέτρου επικάλυψης μεταξύ των προτάσεων δράσης και των σημείων. Οι Behl et al. 2017 ασχολούνται με την ανίχνευση και τον εντοπισμό ενεργειών σε πραγματικό χρόνο μέσω της λήψης προτάσεων δράσης ανά καρέ και την πρόταση ενός αλγορίθμου σύνδεσης που είναι σε θέση να κατασκευάσει και να ενημερώνει τα action tubes ανά καρέ. Πιο πρόσφατα, οι Soomro and Shah 2017 προσπάθησαν να ασχοληθούν με το πρόβλημα της ανίχνευσης και τον εντοπισμό δράσης χωρίς επίβλεψη. Η προσέγγισή τους περιελάμβανε αρχικά την εξόρυξη κατακερματισμένων supervoxel και στη συνέχεια την ανάθεση ένα βάρους σε κάθε supervoxel. Με την εξαγωγή supervoxels, δημιουργούν ένα γράφημα και στη συνέχεια χρησιμοποιούν μια διακριτική clustering προσέγγιση εκπαιδεύεται ένας ταξινομητής.

Η εισαγωγή του R-CNN (Girshick et al. 2014) κατάφερε σημαντικές βελτιώσεις στην απόδοση των δικτύων εντοπισμού αντικειμένων. Αυτή η αρχιτεκτονική, πρώτον, προτείνει περιοχές στην εικόνα που είναι πιθανό να περιέχουν κάποιο αντικείμενο και στη συνέχεια, τα ταξινομεί χρησιμοποιώντας ένα SVM. Εμπνευσμένοι από αυτή την αρχιτεκτονική, οι Gkioxari and Malik 2015 σχεδιάσαν ένα δίκτυο RCNN 2-stream για να προτείνει προτάσεις δράσεων για κάθε καρέ, ένα stream για το επίπεδο καρέ και ένα για την οπτική ροή. Στη συνέχεια, τα συνδέουν χρησιμοποιώντας τον αλγόριθμο σύνδεσης Viterbi. Οι Weinzaepfel, Harchaoui, and Schmid 2015 επεκτείνουν αυτή την προσέγγιση, εκτελώντας προτάσεις στο επίπεδο καρέ και χρησιμοποιώντας ένα τρακερ για τη σύνδεση των προτάσεων αυτών χαρακτηριστικά της χωρικής και οπτικής ροής. Επίσης, η μέθοδός τους εκτελεί χρονικό εντοπισμό μέσω της χρήσης ενός συρόμενου παράθυρου πάνω από τα εντοπισμένα tubes.

Η εισαγωγή του Faster RCNN (Ren et al. 2017) συνήσφερε πολύ τη βελτίωση της απόδοσης των δικτύων εντοπισμού δράσης. Οι Peng and Schmid 2016 και Saha et al. 2016 χρησιμοποιούν το Faster R-CNN αντί για το RCNN για προτάσεις σε επίπεδο καρέ, χρησιμοποιώντας το RPN για εικόνες RGB και οπτικής ροής. Αφού λάβουν χωρικές προτάσεις και προτάσεις κίνησης, οι Peng and Schmid 2016 τις συγχωνεύουν και από κάθε προτεινόμενη ROI, παράγουν 4 ROIs για να επικεντρωθούν σε συγκεκριμένο μέρος του σώματος του δρώντα. Μετά από αυτό, συνδέουν την πρόταση χρησιμοποιώντας τον αλγόριθμο Viterbi για κάθε κλάση και εκτελούν χρονικό εντοπισμό χρησιμοποιώντας ένα συρόμενο παράθυρο,

με πολλαπλές χρονικές κλίμακες και διασκελισμό κάνοντας χρήση μιας μεθόδου μέγιστης υποσυστοιχίας. Απ΄ την άλλη, οι Saha et al. 2016 εκτελούν, επίσης, ταξινόμηση σε επίπεδο καρέ. Μετά απ΄ αυτό, η μέθοδός τους εκτελεί σύντηξη με βάση έναν συνδυασμό της εμφάνισης και της κίνησης με βάση τις προτάσεις και την βαθμολογία αλληλεπικάλυψης. Τέλος, η χρονική προσαρμογή λαμβάνει χώρα χρησιμοποιώντας δυναμικό προγραμματισμό. Παράλληλα, οι Weinzaepfel, Martin, and Schmid 2016 χρησιμοποιούν το Faster RCNN για την εξαγωγή ανθρώπινων tubes από βίντεο που εστιάζουν στο πρόβλημα του ασθενώς εποπτευόμενου εντοπισμού δράσης. Στη συνέχεια, χρησιμοποιώντας πυκνές τροχιές και μια multi-fold Multiple Instance Learning προσέγγιση (Cinbis, Verbeek, and Schmid 2016) εκπαιδεύουν ένα ταξινομητή. Οι Mettes and Snoek 2017 εισήγαγαν μια μέθοδο για zero-shot Εντοπισμού δράσης. Η προσέγγισή τους περιλαμβάνει την βαθμολόγηση των προτεινόμενων action tubes σύμφωνα με τις αλληλεπιδράσεις μεταξύ των ατόμων που δρουν και αντικειμένων. Χρησιμοποίησαν το Faster-RCNN, στο πρώτο βήμα, για την ανίχνευση τόσο των ανθρώπων που δρουν όσο και των αντικειμένων και μετά, χρησιμοποιώντας χωρικές σχέσεις μεταξύ τους, συνδέουν τα προτεινόμενα πλαίσια στον άξονα του χρόνου βασιζόμενοι στην zero-shot πιθανότητα της παρουσίας των ατόμων, συναφών αντικειμένων γύρω απ΄ αυτούς και τις αναμενόμενες χωρικές σχέσεις μεταξύ αντικειμένων και ανθρώπων που δρουν. Επιπλέον οι He et al. 2018 πρότειναν το Tube Proposal Network (TPN) για τη δημιουργία ανεξαρτήτου κλάσης προτάσεις tubelet, οι οποίες χρησιμοποιούν το Faster R-CNN για να λάβουν δισδιάστατες προτάσεις περιοχών και έναν αλγόριθμο σύνδεση για τη σύνδεση των tubelets με τις προτάσεις των περιοχών. Πιο πρόσφατα, οι Girdhar et al. 2018 πρότειναν μια μέθοδο για εντοπισμό δράσεων στο σύνολο δεδομένων AVA (Gu et al. 2018) συνδυάζοντας τις αρχιτεκτονικές των I3D (Carreira and Zisserman 2017) και Faster RCNN. Χρησιμοποιούν μπλοκ του I3D για την λήψη αναπαράστασης βίντεο και το RPN του Faster-RCNN για να προτείνει προτάσεις «ανθρώπου» για το κεντρικό πλαίσιο.

Παράλληλα μ΄ αυτά, οι **singh2016online** και **kalogeiton17iccv:hal-01519812** σχεδίασαν τα δίκτυα τους με βάση το Single Shot Multibox Detector **DBLP:journals/corr/LiuAESR15**). Οι **singh2016online** δημιούργησαν ένα χωροχρονικό δίκτυο πραγματικού χρόνου. Για να λειτουγεί το δίκτυο τους σε πραγματικού χρόνου, **singh2016online** πρότειναν έναν νέο και αποδοτικό αλγόριθμο με την προσθήκη πλαισίων σε tubes σε κάθε καρέ, εάν επικαλύπτονται περισσότερο από ένα κατώφλιοή, ή εναλλακτικά, τερματίζουν το action tube εάν για k καρέ δεν προσθέθηκε κανένα πλαίσιο. Οι **kalogeiton17iccv:hal-01519812** σχεδίασαν ένα δίκτυο δύο ροών, το οποίο κάλεσαν ACT-detector , και εισήγαγαν τα κυβικά (cuboids) anchors. Για K καρέ, και για τα δύο δίκτυα, οι **kalogeiton17iccv:hal-01519812** εξάγουν χωρικά χαρακτηριστικά σε επίπεδο καρέ, στη συνέχεια, τα στοιβάζουν. Τέλος, με τη χρήση των κυβικών anchors, το δίκτυο εξάγει tubelets, με τις αντίστοιχες βαθμολογίες κατάταξης και στόχους παλινδρόμησης. Για τη σύνδεση των τουμπέλετς, οι **kalogeiton17iccv:hal-01519812** ακολουθούν τα ίδια βήματα με τους **singh2016online**. Για χρονική εντοπισμό, χρησιμοποιούν μιά προσέγγιση χρονικής εξομάλυνσης.

Πιο πρόσφατα, το δίκτυο YOLO (**DBLP:journals/corr/RedmonDGF15**) έγινε η έμπνευση για τους **DBLP:journals/corr/abs-1903-00304** και τους

DBLP:journals/corr/abs-1802-08362. Στην προσέγγιση που προτάθηκε από τους **DBLP:journals/corr/abs-1903-00304**, οι έννοιες της εξέλιξης και τού ποσοστού προόδου εισήχθησαν. Εκτός από την πρόταση πλαισίων οριοθέτησης σε επίπεδο καρέ, χρησιμοποιούν το YOLO μαζί με έναν ταξινομητή RNN για να εξάγουν χρονικές πληροφορίες για τις προτάσεις. Με βάση αυτές τις πληροφορίες, δημιουργούν action tubes, χωρίζοντας τα σε κλάσεις. Ορισμένες άλλες προσεγγίσεις περιλαμβάνουν εκτίμηση πόζας αυτή των **DBLP:journals/corr/abs-1802-09232**. Πρότειναν μια μέθοδο υπολογισμού των δισδιάστατων και τρισδιάστατων πόζων και στη συνέχεια εκτέλεσαν ταξινόμηση δράσεων. Χρησιμοποιούν το διαφορίσιμο Soft-argamax για την εκτίμηση των 2D και 3D αρθρώσεων, επειδή η συναρτηση argmax δεν είναι διαφορίσιμη. Στη συνέχεια, για T παρακείμενες δημιουργούν μια απεικόνιση εικόνας με το χρόνο και τις N_j αρθρώσεις ως $x - y$ άξονες, έχοντας 2 κανάλια για την 2D πόζα και 3 για την 3D πόζα. Χρησιμοποιούν Convolutional Layers για να παράγουν χάρτες θερμότητες δράσης και στη συνέχεια χρησιμοποιώντας max plus min pooling και την συνάρτηση softmax εκτελούν ταξινόμηση δράσης. Οι **DBLP:journals/corr/ZolfaghariOSB17** πρότειναν μια αρχιτεκτονική τριών ροών που περιλαμβάνει 2D πόζα, οπτική ροή και πληροφορίες RGB. Αυτά τα streams ενώνονται μέσω του μοντέλου της αλυσίδας Μάρκοφ. Επιπλέον, οι **8237881** πρότειναν μια αρχιτεκτονική με τη χρήση ενός χρονικού convolutional δικτύου παλινδρόμησης, για να πιάνουν την μακροπρόθεσμη εξάρτηση και πληροφίες μεταξύ γειτονικών καρέ και ένα χωρικό δίκτυο παλινδρόμησης, για προτάσεις ανά καρέ. Χρησιμοποιούν μεθόδους παρακολούθησης και δυναμικού προγραμματισμού για τη δημιουργία προτάσεων δράσης.

Τα περισσότερα από τα προαναφερθέντα δίκτυα χρησιμοποιούν ανά καρέ χωρικές προτάσεις και εξάγουν τις χρονικές τους πληροφορίες υπολογίζοντας την οπτική ροή. Από την άλλη οι **DBLP:journals/corr/SahaSC17** σχεδίασαν μια αρχιτεκτονική η οποία περιλαμβάνει προτάσεις σε επίπεδο τμήματος βίντεο, το οποίο σημαίνει περισσότερα από ένα καρέ ταυτόχρονα. Οι **DBLP:journals/corr/SahaSC17** πρότειναν μια 3D-RPN αρχιτεκτονική που είναι σε θέση να δημιουργήσει και να ταξινομήσει τρισδιάστες προτάσεις αποτελούμενες από 2 συνεχόμενα καρέ. Επίσης, πρότειναν έναν αλγόριθμο σύνδεσης, τροποποιώντας αυτόν που πρότειναν οι Saha et al. 2016. Πάνω σ΄ αυτό, οι **DBLP:journals/corr/HouCS17** σχεδιάσαν μια αρχιτεκτονική για τη δημιουργία προτάσεων δράσης για περισσότερα από 2 καρέ, καλώντας το μοντέλο τους Tube CNN (T-CNN). Στην προσέγγισή τους, το επίπεδο του τμήματος βίντεο σημαίνει ότι ολόκληρο το βίντεο χωρίζεται κλιπ βίντεο ίδιου αριθμού καρέ και με τη χρήση του C3D για την εξόρυξη χαρακτηριστικών, επιστρέφουν χωροχρονικές προτάσεις. Μετά την λήψη των προτάσων, οι **ΔΒΛΠ:θουρναλς/ςορρ/Ηου"Σ17** συνδέουν τις tube προτάσεις τους με έναν αλγόριθμο στηριζόμενος στην πιθανότητα ύπαρξης δράσης και την επικάλυψη μεταξύ των tubes. Τέλος, η λειτουργία ταξινόμησης λαμβάνει χώρα για τα συνδεδεμένα action tubes.

Βιβλιογραφία

[1] Ο. ἢοματ και Θ. Λ. ̃ροωλεψ. 'Προβαβιλιστις ρεϛογνιτιον οφ αϛτιιτψ υσινγ λοϛαλ αππεαρανϛε'. Στο: *Προϛεεδινγϛ. 1999 ΙΕΕΕ ̃δμπυτερ Σοϛιετψ ̃δνφερενϛε ον ̃δμπυτερ ΐσιον ανδ Παττερν Ρεϛογνιτιον (̃ατ. Νο ΠΡ00149)*. Τόμ. 2. 1999, 104–109 ̃όλ. 2. ΔΟΙ: 10.1109/˜˜ΠΡ.1999.784616.

[2] Α. Φ. Βοβιϛκ και Θ. Ω. Δαις. 'Τηε ρεϛογνιτιον οφ ηυμαν μοεμεντ υσινγ τεμποραλ τεμπλατες'. Στο: *ΙΕΕΕ Τρανϛαϛτιονϛ ον Παττερν Αναλψϛις ανδ Μαϛηινε Ιντελλιγενϛε* 23.3 (2001), σσ. 257–267. ΔΟΙ: 10.1109/34.910878.

[3] Λ. Ζελνικ-Μανορ και Μ. Ιρανι. 'Εεντ-βασεδ αναλψϛις οφ ιδεο'. Στο: *Προϛεεδινγϛ οφ τηε 2001 ΙΕΕΕ ̃δμπυτερ Σοϛιετψ ̃δνφερενϛε ον ̃δμπυτερ ΐσιον ανδ Παττερν Ρεϛογνιτιον. ˜ΠΡ 2001*. Τόμ. 2. 2001, σσ. ΙΙ–ΙΙ. ΔΟΙ: 10.1109/˜˜ΠΡ.2001.990935.

[4] Λαπτε και Λινδεβεργ. 'Σπαϛε-τιμε ιντερεϛτ ποιντϛ'. Στο: *Προϛεεδινγϛ Νιντη ΙΕΕΕ Ιντερνατιοναλ ̃δνφερενϛε ον ̃δμπυτερ ΐσιον*. 2003, 432–439 ολ.1. ΔΟΙ: 10.1109/I˜˜˜.2003.1238378.

[5] Αλπερ Ψιλμαζ και Μυβαρακ Σηαη. 'Αϛτιονϛ σκετϛη: α νοελ αϛτιον ρεπρεϛεντατιον'. Στο: *2005 ΙΕΕΕ ̃δμπυτερ Σοϛιετψ ̃δνφερενϛε ον ̃δμπυτερ ΐσιον ανδ Παττερν Ρεϛογνιτιον (˜ΠΡ'05)*. Τόμ. 1. 2005, 984–989 ολ. 1. ΔΟΙ: 10.1109/˜˜ΠΡ.2005.58.

[6] Μ. Βλανκ κ.ά. 'Αϛτιονϛ ας σπαϛε-τιμε σηαπεσ'. Στο: *Τεντη ΙΕΕΕ Ιντερνατιοναλ ̃δνφερενϛε ον ̃δμπυτερ ΐσιον (Ι˜˜˜'05) ̃όλυμε 1*. Τόμ. 2. 2005, 1395–1402 ̃όλ. 2. ΔΟΙ: 10.1109/I˜˜˜.2005.28.

[7] Π. Δολλαρ κ.ά. 'Βεηαιορ ρεϛογνιτιον ια σπαρϛε σπατιο-τεμποραλ φεατυρεσ'. Στο: *2005 ΙΕΕΕ Ιντερνατιοναλ Ωορκϛηοπ ον ΐσυαλ Συρϛειλλανϛε ανδ Περφορμανϛε Εαλυατιον οφ Τραϛκινγ ανδ Συρϛειλλανϛε*. 2005, σσ. 65–72. ΔΟΙ: 10.1109/˜ΣΠΕΤΣ.2005.1570899.

[8] Ε. Σηεϛητμαν και Μ. Ιρανι. 'Σπαϛε-τιμε βεηαιορ βασεδ ϛορρελατιον'. Στο: *2005 ΙΕΕΕ ̃δμπυτερ Σοϛιετψ ̃δνφερενϛε ον ̃δμπυτερ ΐσιον ανδ Παττερν Ρεϛογνιτιον (˜ΠΡ'05)*. Τόμ. 1. 2005, 405–412 ολ. 1. ΔΟΙ: 10.1109/˜˜ΠΡ.2005.328.

[9] Ψ. Σηεικη, Μ. Σηεικη και Μ. Σηαη. 'Εξπλορινγ τηε σπαϛε οφ α ηυμαν αϛτιον'. Στο: *Τεντη ΙΕΕΕ Ιντερνατιοναλ ̃δνφερενϛε ον ̃δμπυτερ ΐσιον (Ι˜˜˜'05) ̃όλυμε 1*. Τόμ. 1. 2005, 144–149 ̃όλ. 1. ΔΟΙ: 10.1109/I˜˜˜.2005.90.

[10] Α. Ψιλμαζ και Μ. Σηαη. 'Ρεcογνιζινγ ηυμαν αcτιονς ιν ιδεος ασχυιρεδ βψ υνcαλιβρατεδ μοινγ cαμερας'. Στο: *Τεντη ΙΕΕΕ Ιντερνατιοναλ δνφερενcε ον δμπυτερ ϊσιον (Ι˜˜05) ολυμε 1*. Τόμ. 1. 2005, 150–157 ŏλ. 1. DOI: 10.1109/I˜˜˜.2005.201.

[11] Θυαν ἁρλος Νιεβλες, Ηονγcηενγ Ωανγ και Φει Φει Λι. 'Υνσυπερισεδ Λεαρνινγ οφ Ηυμαν Αcτιον ἁτεγοριες Υσινγ Σπατιαλ-Τεμποραλ Ωορδς.' Στο: τόμ. 79. Σεπτ. 2006, σσ. 1249–1258.

[12] Μ. Σ. Ρψοο και Θ. Κ. Αγγαρωαλ. 'Σεμαντις Υνδερcτανδινγ οφ ὀντινυεδ ανδ Ρεcυρcιε Ηυμαν Αcτιτιες'. Στο: *18τη Ιντερνατιοναλ ὀνφερενcε ον Παττερν Ρεcογνιτιον (ΓΠΡ'06)*. Τόμ. 1. 2006, σσ. 379–378. DOI: 10.1109/IˆΠΡ.2006.1043.

[13] Ψ. Κε, Ρ. Συκτηανκαρ και Μ. Ηεβερτ. 'Σπατιο-τεμποραλ Σηαπε ανδ Φλοω ὀρρελατιον φορ Αcτιον Ρεcογνιτιον'. Στο: *2007 ΙΕΕΕ ὀνφερενcε ον ὀμπυτερ ϊσιον ανδ Παττερν Ρεcογνιτιον*. 2007, σσ. 1–8. DOI: 10.1109/ˆ˜ΠΡ.2007.383512.

[14] Μ. Δ. Ροδριγυεζ, Θ. Αημεδ και Μ. Σηαη. 'Αcτιον ΜΑˆΗ α σπατιο-τεμποραλ Μαξιμυμ Αεραγε ὀρρελατιον Ηειγητ φιλτερ φορ αcτιον ρεcογνιτιον'. Στο: *2008 ΙΕΕΕ ὀνφερενcε ον ὀμπυτερ ϊσιον ανδ Παττερν Ρεcογνιτιον*. 2008, σσ. 1–8. DOI: 10.1109/ˆ˜ΠΡ.2008.4587727.

[15] Π. Φ. Φελζενσζωαλβ κ.ά. 'Οβθεcτ Δετεcτιον ωιτη Δισcριμινατιελψ Τραινεδ Παρτ-Βασεδ Μοδελς'. Στο: *ΙΕΕΕ Τρανσαcτιονς ον Παττερν Αναλψσις ανδ Μαcηινε Ιντελλιγενcε* 32.9 (2010), σσ. 1627–1645. DOI: 10.1109/TΠΑΜΙ.2009.167.

[16] Θ.Κ. Αγγαρωαλ και Μ.Σ. Ρψοο. 'Ηυμαν Αcτιτψ Αναλψσις: Α Ρειεω'. Στο: *Aˆ˜M ὀμπυτ. Συρ.* 43.3 (Απρ. 2011), 16:1–16:43. ISSN: 0360-0300. DOI: 10.1145/1922649.1922653. URL: ηττπ://δοι.αcμ.οργ/10.1145/1922649.1922653.

[17] Σ. Θι κ.ά. '3Δ ὀνολυτιοναλ Νευραλ Νετωορκς φορ Ηυμαν Αcτιον Ρεcογνιτιον'. Στο: *ΙΕΕΕ Τρανσαcτιονς ον Παττερν Αναλψσις ανδ Μαcηινε Ιντελλιγενcε* 35.1 (2013), σσ. 221–231. DOI: 10.1109/TΠΑΜΙ.2012.59.

[18] Σ. Μανεν, Μ. Γυιλλαυμιν και Λ. ˝. Γοολ. 'Πριμε Οβθεcτ Προποcαλς ωιτη Ρανδομιζεδ Πριμ΄ς Αλγοριτημ'. Στο: *2013 ΙΕΕΕ Ιντερνατιοναλ ὀνφερενcε ον ὀμπυτερ ϊσιον*. 2013, σσ. 2536–2543. DOI: 10.1109/Iˆ˜˜.2013.315.

[19] Ψ. Τιαν, Ρ. Συκτηανκαρ και Μ. Σηαη. 'Σπατιοτεμποραλ Δεφορμαβλε Παρτ Μοδελς φορ Αcτιον Δετεcτιον'. Στο: *2013 ΙΕΕΕ ὀνφερενcε ον ὀμπυτερ ϊσιον ανδ Παττερν Ρεcογνιτιον*. 2013, σσ. 2642–2649. DOI: 10.1109/ˆ˜ΠΡ.2013.341.

[20] Θ.Ρ.Ρ. Υιθλινγς κ.ά. 'Σελεcτιε Σεαρcη φορ Οβθεcτ Ρεcογνιτιον'. Στο: *Ιντερνατιοναλ Θουρναλ οφ ὀμπυτερ ϊσιον* (2013). DOI: 10.1007/σ11263-013-0620-5. URL: ηττπ://ωωω.ηυππελεν.νλ/πυβλιcατιονς/σελεcτιεΣεαρcηΔραφτ.πδφ.

[21] Ρ. Γιρσηιςκ κ.ά. 'Ριςη Φεατυρε Ηιεραρςηιες φορ Αςςυρατε Οβϑεςτ Δετεςτιον ανδ Σεμαντις Σεγμεντατιον'. Στο: *2014 ΙΕΕΕ ὀνφερενςε ον ὀμπυτερ ἴσιον ανδ Παττερν Ρεςογνιτιον*. 2014, σσ. 580–587. ΔΟΙ: 10.1109/ˆˇΠΡ.2014.81.

[22] Μ. Θαιν κ.ά. 'Αςτιον Λοςαλιζατιον ωιτη Τυβελετς φρομ Μοτιον'. Στο: *2014 ΙΕΕΕ ὀνφερενςε ον ὀμπυτερ ἴσιον ανδ Παττερν Ρεςογνιτιον*. 2014, σσ. 740–747. ΔΟΙ: 10.1109/ˆˇΠΡ.2014.100.

[23] Α. Καρπατηψ κ.ά. 'Λαργε-Σςαλε ἴδεο ˆλασσιφιςατιον ωιτη ὀνολυτιοναλ Νευραλ Νετωορκσ'. Στο: *2014 ΙΕΕΕ ὀνφερενςε ον ὀμπυτερ ἴσιον ανδ Παττερν Ρεςογνιτιον*. 2014, σσ. 1725–1732. ΔΟΙ: 10.1109/ˆˇΠΡ.2014.223.

[24] Δαν Ονεατα κ.ά. 'Σπατιο-Τεμποραλ Οβϑεςτ Δετεςτιον Προποσαλς'. Στο: Σεπτ. 2014. ΔΟΙ: 10.1007/978-3-319-10578-9ˆ48.

[25] Καρεν Σιμονψαν και Ανδρεω Ζισσερμαν. 'Τωο-στρεαμ ςονολυτιοναλ νετωορκς φορ αςτιον ρεςογνιτιον ιν ιδεοσ'. Στο: *Αδανςες ιν Νευραλ Ινφορματιον Προςεσσινγ Σψστεμς*. 2014, σσ. 568–576.

[26] Ω. ἦεν και Θ. Θ. ὀρσο. 'Αςτιον Δετεςτιον βψ Ιμπλιςιτ Ιντεντιοναλ Μοτιον ˆλυστερινγ'. Στο: *2015 ΙΕΕΕ Ιντερνατιοναλ ὀνφερενςε ον ὀμπυτερ ἴσιον (Ι˜˜)*. 2015, σσ. 3298–3306. ΔΟΙ: 10.1109/Iˆˆˇ.2015.377.

[27] Θαν ˆ. αν Γεμερτ κ.ά. 'ΑΠΤ: Αςτιον λοςαλιζατιον προποσαλς φρομ δενσε τραϑεςτοριες'. Στο: *Προςεεδινγς οφ τηε Βριτιση Μαςηινε ἴσιον ὀνφερενςε (ΒΜ˜)*. ΒΜ˜Α Πρεσς, 2015, σσ. 177.1–177.12. ΙΣΒΝ: 1-901725-53-7. ΔΟΙ: 10.5244/ˆ.29.177. ΥΡΛ: ητττπς://δξ.δοι.οργ/10.5244/ˆ.29.177.

[28] Γ. Γκιοξαρι και Θ. Μαλικ. 'Φινδινγ αςτιον τυβες'. Στο: *2015 ΙΕΕΕ ὀνφερενςε ον ὀμπυτερ ἴσιον ανδ Παττερν Ρεςογνιτιον (˜ΠΡ)*. 2015, σσ. 759–768. ΔΟΙ: 10.1109/ˆˇΠΡ.2015.7298676.

[29] Θοε Ψυε-Ηει Νγ κ.ά. 'Βεψονδ σηορτ σνιππετς: Δεεπ νετωορκς φορ ιδεο ςλασσιφιςατιον'. Στο: *2015 ΙΕΕΕ ὀνφερενςε ον ὀμπυτερ ἴσιον ανδ Παττερν Ρεςογνιτιον (˜ΠΡ)*. 2015, σσ. 4694–4702. ΔΟΙ: 10.1109/ˆˇΠΡ.2015.7299101.

[30] Κ. Σοομρο, Η. Ιδρεες και Μ. Σηαη. 'Αςτιον Λοςαλιζατιον ιν ἴδεος τηρουγη ὀντεξτ Ωαλκ'. Στο: *2015 ΙΕΕΕ Ιντερνατιοναλ ὀνφερενςε ον ὀμπυτερ ἴσιον (Ι˜˜)*. 2015, σσ. 3280–3288. ΔΟΙ: 10.1109/Iˆˆˇ.2015.375.

[31] Δ. Τραν κ.ά. 'Λεαρνινγ Σπατιοτεμποραλ Φεατυρες ωιτη 3Δ ὀνολυτιοναλ Νετωορκς'. Στο: *2015 ΙΕΕΕ Ιντερνατιοναλ ὀνφερενςε ον ὀμπυτερ ἴσιον (Ι˜˜)*. 2015, σσ. 4489–4497. ΔΟΙ: 10.1109/Iˆˆˇ.2015.510.

[32] Π. Ωεινζαεπφελ, Ζ. Ηαρσηαουι και ˆ. Σχημιδ. 'Λεαρνινγ το Τραςκ φορ Σπατιο-Τεμποραλ Αςτιον Λοςαλιζατιον'. Στο: *2015 ΙΕΕΕ Ιντερνατιοναλ ὀνφερενςε ον ὀμπυτερ ἴσιον (Ι˜˜)*. 2015, σσ. 3164–3172. ΔΟΙ: 10.1109/Iˆˆˇ.2015.362.

[33] Γ. Ψυ και Θ. Ψυαν. 'Φαστ αςτιον προποσαλς φορ ηυμαν αςτιον δετεςτιον ανδ σεαρςη'. Στο: *2015 ΙΕΕΕ ὀνφερενςε ον ὀμπυτερ ἴσιον ανδ Παττερν Ρεςογνιτιον*. 2015, σσ. 1302–1311. ΔΟΙ: 10.1109/ˆˇΠΡ.2015.7298735.

[34] Ρ. Γ. ῖνβις, Θ. ἔρβεεκ και ῎. Σχημιδ. ‘Ωεακλψ Συπερισεδ Οβθεστ Λοχαλιζατιον ωιτη Μυλτι-Φολδ Μυλτιπλε Ινστανχε Λεαρνινγ’. Στο: *IEEE Τρανσαχτιονς ον Παττερν Αναλψσις ανδ Μαχηινε Ιντελλιγενχε* 39.1 (2016), σσ. 189–203. ΔΟΙ: 10.1109/ΤΠΑΜΙ.2016.2535231.

[35] ῎. Φειχητενηοφερ, Α. Πινζ και Α. Ζισσερμαν. ‘ὃνολυτιοναλ Τωο-Στρεαμ Νετωορκ Φυσιον φορ ΐδεο Αχτιον Ρεχογνιτιον’. Στο: *2016 IEEE ὃνφερενχε ον ὃμπυτερ ΐσιον ανδ Παττερν Ρεχογνιτιον (῍ΠΡ)*. 2016, σσ. 1933–1941. ΔΟΙ: 10.1109/῍῍ΠΡ.2016.213.

[36] Κ. Ηε κ.ά. ‘Δεεπ Ρεσιδυαλ Λεαρνινγ φορ Ιμαγε Ρεχογνιτιον’. Στο: *2016 IEEE ὃνφερενχε ον ὃμπυτερ ΐσιον ανδ Παττερν Ρεχογνιτιον (῍ΠΡ)*. 2016, σσ. 770–778. ΔΟΙ: 10.1109/῍῍ΠΡ.2016.90.

[37] Πασχαλ Μεττες, Θαν ῎. αν Γεμερτ και ἓες Γ. Μ. Σνοεκ. ‘Σποτ Ον: Αχτιον Λοχαλιζατιον φρομ Ποιντλψ-Συπερισεδ Προποσαλσ’. Στο: *ὃPP* αβς/1604.07602 (2016). αρΞι: 1604.07602. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1604.07602.

[38] Ξιαοθιανγ Πενγ και ὃρδελια Σχημιδ. ‘Μυλτι-ρεγιον τωο-στρεαμ Ρ-῎ΝΝ φορ αχτιον δετεχτιον’. Στο: *Ε῍῍῍- Ευροπεαν ὃνφερενχε ον ὃμπυτερ ΐσιον*. Τόμ. 9908. Λεχτυρε Νοτες ιν ὃμπυτερ Σχιενχε. Αμστερδαμ, Νετηερλανδς: Σπρινγερ, Οκτ. 2016, σσ. 744–759. ΔΟΙ: 10.1007/978-3-319-46493-0῎῍45. ΥΡΛ: ηττπς://ηαλ.ινρια.φρ/ηαλ-01349107.

[39] Συμαν Σαηα κ.ά. ‘Δεεπ Λεαρνινγ φορ Δετεχτινγ Μυλτιπλε Σπαχε-Τιμε Αχτιον Τυβες ιν ΐδεοσ’. Στο: *ὃPP* αβς/1608.01529 (2016). αρΞι: 1608.01529. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1608.01529.

[40] Λιμιν Ωανγ κ.ά. ‘Τεμποραλ Σεγμεντ Νετωορκς: Τοωαρδς Γοοδ Πραχτιχες φορ Δεεπ Αχτιον Ρεχογνιτιον’. Στο: *ὃPP* αβς/1608.00859 (2016). αρΞι: 1608.00859. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1608.00859.

[41] Πηιλιππε Ωεινζαεπφελ, Ξαιερ Μαρτιν και ὃρδελια Σχημιδ. ‘Τοωαρδς Ωεακλψ-Συπερισεδ Αχτιον Λοχαλιζατιον’. Στο: *ὃPP* αβς/1605.05197 (2016). αρΞι: 1605.05197. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1605.05197.

[42] Βοωεν Ζηανγ κ.ά. ‘Ρεαλ-τιμε Αχτιον Ρεχογνιτιον ωιτη Ενηανχεδ Μοτιον ἔχτορ ῎ΝΝς’. Στο: *ὃPP* αβς/1604.07669 (2016). αρΞι: 1604.07669. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1604.07669.

[43] Ηαρκιρατ Σ. Βεηλ κ.ά. ‘Ινχρεμενταλ Τυβε ὃνστρυχτιον φορ Ηυμαν Αχτιον Δετεχτιον’. Στο: *ὃPP* αβς/1704.01358 (2017). αρΞι: 1704.01358. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1704.01358.

[44] Θ. ἂρρειρα και Α. Ζισσερμαν. ‘Χυο ἄδις, Αχτιον Ρεχογνιτιον; Α Νεω Μοδελ ανδ τηε Κινετιχς Δατασετ’. Στο: *2017 IEEE ὃνφερενχε ον ὃμπυτερ ΐσιον ανδ Παττερν Ρεχογνιτιον (῍ΠΡ)*. 2017, σσ. 4724–4733. ΔΟΙ: 10.1109/῍῍ΠΡ.2017.502.

[45] Αλι Διβα κ.ά. ‘Τεμποραλ 3Δ ὃνΝετς: Νεω Αρχηιτεχτυρε ανδ Τρανσφερ Λεαρνινγ φορ ΐδεο ῏λασσιφιχατιον’. Στο: *ὃPP* αβς/1711.08200 (2017). αρΞι: 1711.08200. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1711.08200.

[46] Θ. Δοναηυε κ.ά. 'Λονγ-Τερμ Ρεςυρρεντ δνολυτιοναλ Νετωορκς φορ ϊσυαλ Ρεςογνιτιον ανδ Δεσςριπτιον'. Στο: *IEEE Τρανσαςτιονς ον Παττερν Αναλψσις ανδ Μαςηινε Ιντελλιγενςε* 39.4 (2017), σσ. 677–691. ΔΟΙ: 10.1109/ΤΠΑΜΙ.2016.2599174.

[47] Ροηιτ Γιρδηαρ και Δεα Ραμαναν. 'Αττεντιοναλ Ποολινγ φορ Αςτιον Ρεςογνιτιον'. Στο: *δPP* αβς/1711.01467 (2017). αρΞι: 1711.01467. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1711.01467.

[48] Κ. Ηαρα, Η. Καταοκα και Ψ. Σατοη. 'Λεαρνινγ Σπατιο-Τεμποραλ Φεατυρες ωιτη 3Δ Ρεσιδυαλ Νετωορκς φορ Αςτιον Ρεςογνιτιον'. Στο: *2017 IEEE Ιντερνατιοναλ δνφερενςε ον δμπυτερ ϊσιον Ωορκσηοπς (Ι˜˜Ω)*. 2017, σσ. 3154–3160. ΔΟΙ: 10.1109/Ι˜˜˜Ω.2017.373.

[49] Γ. Ηυανγ κ.ά. 'Δενσελψ δννεςτεδ δνολυτιοναλ Νετωορκς'. Στο: *2017 IEEE δνφερενςε ον δμπυτερ ϊσιον ανδ Παττερν Ρεςογνιτιον (˜ΠΡ)*. 2017, σσ. 2261–2269. ΔΟΙ: 10.1109/˜˜ΠΡ.2017.243.

[50] ℌιη-Ψαο Μα κ.ά. 'ΤΣ-ΛΣΤΜ ανδ Τεμποραλ-Ινςεπτιον: Εξπλοιτινγ Σπατιοτεμποραλ Δψναμιςς φορ Αςτιτψ Ρεςογνιτιον'. Στο: *δPP* αβς/1703.10667 (2017). αρΞι: 1703.10667. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1703.10667.

[51] Π. Μεττες και ˜. Γ. Μ. Σνοεκ. 'Σπατιαλ-Αωαρε Οβϑεςτ Εμβεδδινγς φορ Ζερο-Σηοτ Λοςαλιζατιον ανδ ˜λασσιφιςατιον οφ Αςτιονς'. Στο: *2017 IEEE Ιντερνατιοναλ δνφερενςε ον δμπυτερ ϊσιον (Ι˜˜)*. 2017, σσ. 4453–4462. ΔΟΙ: 10.1109/Ι˜˜˜.2017.476.

[52] Σ. Ρεν κ.ά. 'Φαστερ Ρ-˜ΝΝ: Τοωαρδς Ρεαλ-Τιμε Οβϑεςτ Δετεςτιον ωιτη Ρεγιον Προποσαλ Νετωορκσ'. Στο: *IEEE Τρανσαςτιονς ον Παττερν Αναλψσις ανδ Μαςηινε Ιντελλιγενςε* 39.6 (2017), σσ. 1137–1149. ΔΟΙ: 10.1109/ΤΠΑΜΙ.2016.2577031.

[53] Κ. Σοομρο και Μ. Σηαη. 'Υνσυπερισεδ Αςτιον Δισςοερψ ανδ Λοςαλιζατιον ιν ϊδεος'. Στο: *2017 IEEE Ιντερνατιοναλ δνφερενςε ον δμπυτερ ϊσιον (Ι˜˜)*. 2017, σσ. 696–705. ΔΟΙ: 10.1109/Ι˜˜˜.2017.82.

[54] Ψι Ζηυ κ.ά. 'Ηιδδεν Τωο-Στρεαμ δνολυτιοναλ Νετωορκς φορ Αςτιον Ρεςογνιτιον'. Στο: *δPP* αβς/1704.00389 (2017). αρΞι: 1704.00389. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1704.00389.

[55] Αλι Διβα κ.ά. 'Τεμποραλ 3Δ δνΝετς Υσινγ Τεμποραλ Τρανσιτιον Λαψερ'. Στο: *2018 IEEE δνφερενςε ον δμπυτερ ϊσιον ανδ Παττερν Ρεςογνιτιον Ωορκσηοπς, ˜ΠΡ Ωορκσηοπς 2018, Σαλτ Λακε ϊτψ, ΥΤ, ΥΣΑ, Θυνε 18-22, 2018*. 2018, σσ. 1117–1121. ΥΡΛ: ηττπ://οπεναccεσς.τηεcφ.cομ/cοντεντ˜˜cπρ˜˜2018˜ὼορκσηοπς/ω19/ητμλ/Διβα˜˜Τεμποραλ˜˜3Δ˜˜δνΝετσ˜˜˜ΠΡ˜˜2018˜˜παπερ.ητμλ.

[56] Ροηιτ Γιρδηαρ κ.ά. 'Α Βεττερ Βασελινε φορ Α˝Α'. Στο: *δPP* αβς/1807.10066 (2018). αρΞι: 1807.10066. ΥΡΛ: ηττπ://αρξι.οργ/αβς/1807.10066.

[57] ˜. Γυ κ.ά. 'Α˝Α: Α ϊδεο Δαταςετ οφ Σπατιο-Τεμποραλλψ Λοςαλιζεδ Ατομις ϊσυαλ Αςτιονσ'. Στο: *2018 IEEE/˜Φ δνφερενςε ον δμπυτερ ϊσιον ανδ Παττερν Ρεςογνιτιον*. 2018, σσ. 6047–6056. ΔΟΙ: 10.1109/˜˜ΠΡ.2018.00633.

[58] Μιχηελλε Γυο κ.ά. 'Νευραλ Γραπη Ματχηινγ Νετωορκς φορ Φεωσηοτ 3Δ Αχτιον Ρεχογνιτιον'. Στο: *Τηє Ευροπεαν ὃνφερενχє ον ὃμπυτερ ἴσιον (Ε͒͒)*. 2018.

[59] Θ. Ηε κ.ά. 'Γενεριχ Τυβελετ Προποσαλς φορ Αχτιον Λοχαλιζατιον'. Στο: *2018 ΙΕΕΕ Ωιντερ ὃνφερενχє ον Αππλιχατιονς οφ ὃμπυτερ ἴσιον (ΩΑ͒͒)*. 2018, σσ. 343–351. ΔΟΙ: 10.1109/ΩΑ͒͑.2018.00044.

[60] Ψυ Κονγ και Ψυν Φυ. 'Ηυμαν Αχτιον Ρεχογνιτιον ανδ Πρεδιχτιον: Α Συρεψ'. Στο: *ὃPP* αβς/1806.11230 (2018). αρΞι: 1806.11230. ΥΡΛ: ητττπ://αρξι.οργ/αβς/1806.11230.

[61] Δ. Τραν κ.ά. 'Α ͒λοσερ Λοοκ ατ Σπατιοτεμποραλ ὃνολυτιονς φορ Αχτιον Ρεχογνιτιον'. Στο: *2018 ΙΕΕΕ/͒Φ ὃνφερενχє ον ὃμπυτερ ἴσιον ανδ Πατ͑τερν Ρεχογνιτιον*. 2018, σσ. 6450–6459. ΔΟΙ: 10.1109/͑͒ΠΡ.2018.00675.