

Αναγνώριση Ανθρώπινης Δραστηριότητας σε Πραγματικό Χρόνο με CNN-LSTM

Κασιωτάκης Ηλίας
Τμήμα Ψηφιακών Συστημάτων
Πανεπιστήμιο Πειραιώς, Αθήνα, Ελλάδα

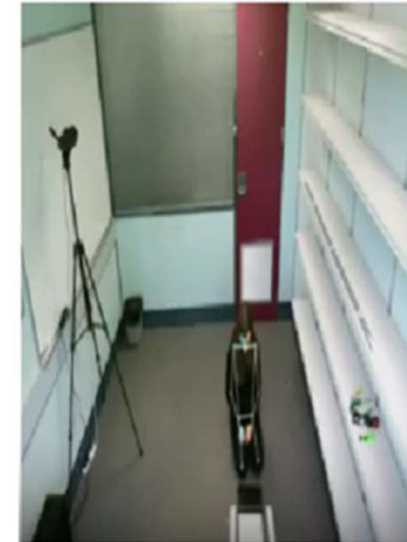
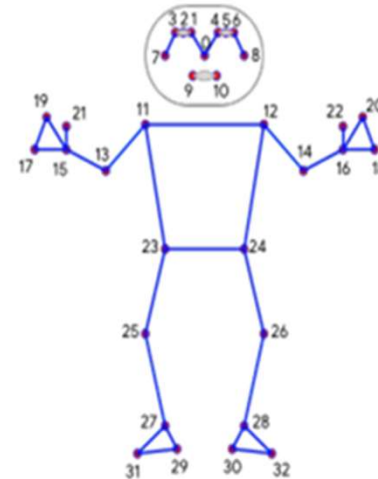
Σκοπός

- Συνεργασία μεταξύ ανθρώπων και συνεργατικών ρομπότ γίνεται σημαντική στη βιομηχανία.
- Χρήση σκελετικών δεδομένων για ανεξαρτησία από το περιβάλλον και object info για εξαγωγή πληροφορίας σχετικά με την αλληλεπίδραση του ανθρώπου με τα αντικείμενα κατά τη δράση.
- Ανάπτυξη CNN-LSTM μοντέλου



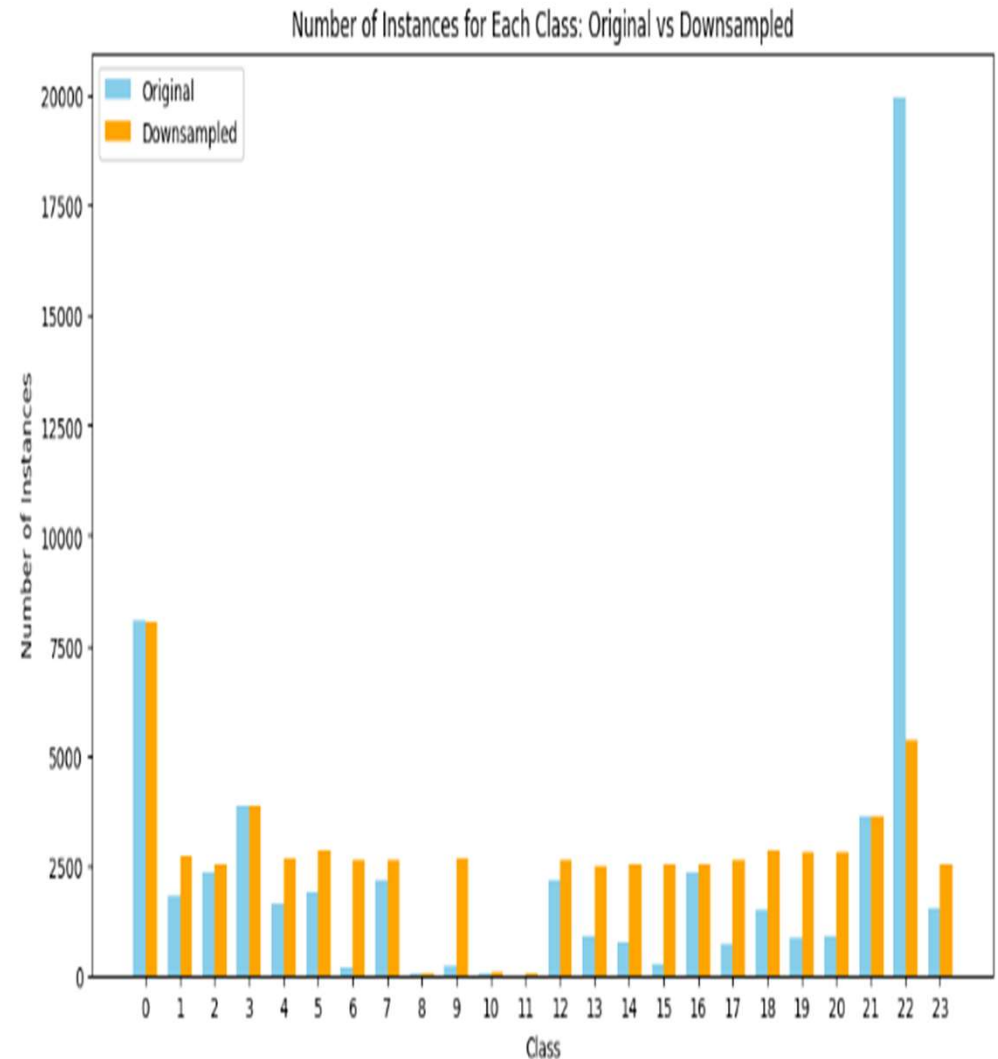
Μεθοδολογία

- Χρήση του συνόλου δεδομένων IKEA ASM για συναρμολόγηση επίπλων.
- Εξαγωγή σκελετικών δεδομένων μέσω του εργαλείου MediaPipe.
- Χρήση κανονικοποιημένων δεδομένων των θέσεων joints και αντικειμένων.
- Αρχιτεκτονική CNN-LSTM για συνδυασμό χωρικών και χρονικών χαρακτηριστικών.



ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ

- Σπάμε video σε μικρότερα των 24 frames
- Εξαγωγή (x,y) joints
- Εισαγωγή (x,y) boxes αντικειμένων (κοντινότερων σε δεξί και αριστερό χέρι (joint 15,16))
- Κανονικοποίηση και μετατόπιση αρχής
- Oversampling-DownSampling

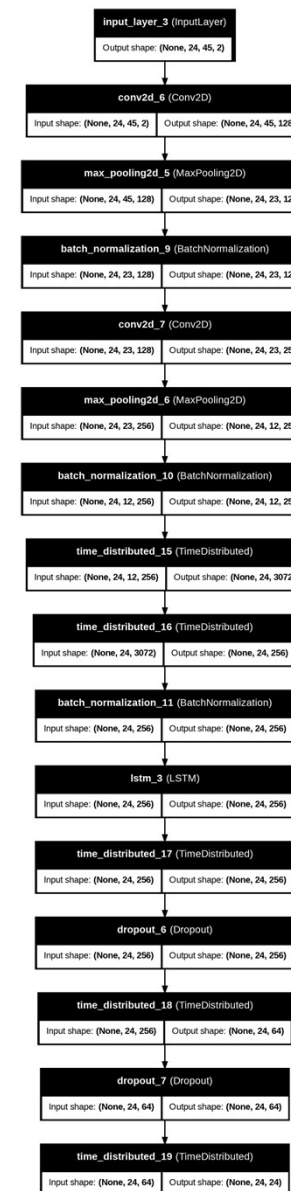


Αρχιτεκτονική του Μοντέλου

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 24, 45, 2)	0
conv2d (Conv2D)	(None, 24, 45, 128)	2,432
conv2d_1 (Conv2D)	(None, 24, 45, 256)	295,168
max_pooling2d (MaxPooling2D)	(None, 24, 23, 256)	0
conv2d_2 (Conv2D)	(None, 24, 23, 256)	590,080
dropout (Dropout)	(None, 24, 23, 256)	0
time_distributed (TimeDistributed)	(None, 24, 5888)	0
time_distributed_1 (TimeDistributed)	(None, 24, 256)	1,507,584
lstm (LSTM)	(None, 24, 256)	525,312
time_distributed_2 (TimeDistributed)	(None, 24, 256)	65,792
dropout_1 (Dropout)	(None, 24, 256)	0
time_distributed_3 (TimeDistributed)	(None, 24, 128)	32,896
dropout_2 (Dropout)	(None, 24, 128)	0
time_distributed_4 (TimeDistributed)	(None, 24, 64)	8,256
dropout_3 (Dropout)	(None, 24, 64)	0
time_distributed_5 (TimeDistributed)	(None, 24, 24)	1,560

Total params: 3,029,080 (11.56 MB)
 Trainable params: 3,029,080 (11.56 MB)
 Non-trainable params: 0 (0.00 B)



Αποτελέσματα



Η ΕΚΠΑΙΔΕΥΣΗ ΤΟΥ
ΜΟΝΤΕΛΟΥ ΣΤΑΜΑΤΗΣΕ
ΛΟΓΩ 'EARLY STOPPING'.

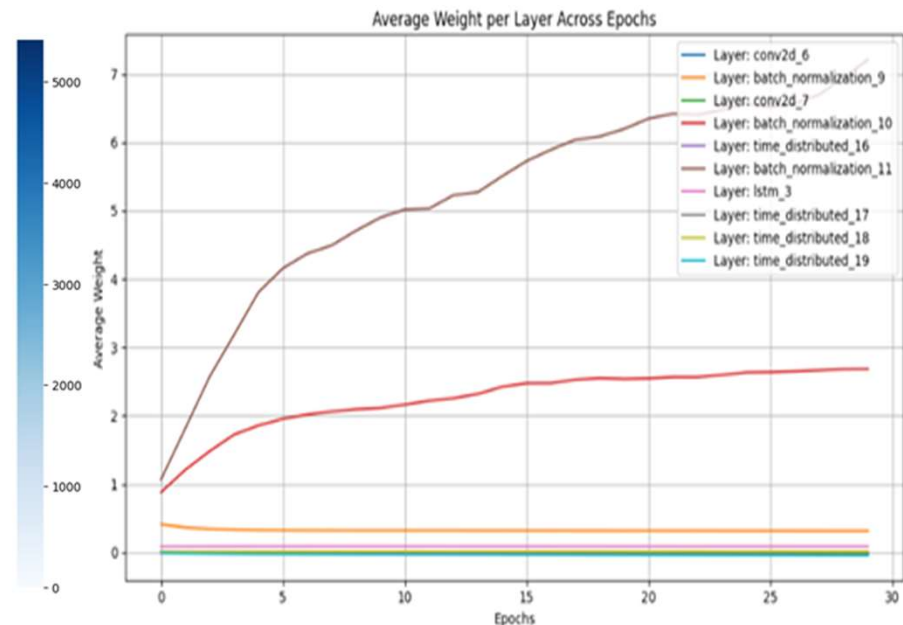
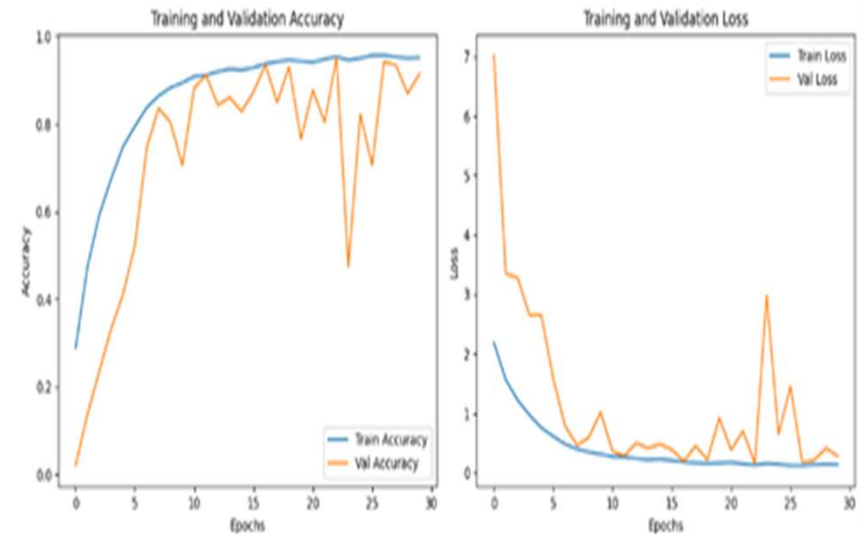


ΤΟ ΜΟΝΤΕΛΟ ΕΙΧΕ ΥΨΗΛΗ
ΑΚΡΙΒΕΙΑ ΣΤΑ ΔΕΔΟΜΕΝΑ
ΕΚΠΑΙΔΕΥΣΗΣ ΚΑΙ
VALIDATION ΑΛΛΑ
ΧΑΜΗΛΗ ΣΤΟ TEST SET.
ΕΠΙΤΥΓΧΑΝΕΙ REAL TIME
(0.04 SEC ΓΙΑ 24 FRAMES)



ΠΡΟΒΛΗΜΑΤΑ
OVERFITTING ΚΑΙ
VANISHING GRADIENT.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	292	0	0	0	137	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	556	289	12	247	0	0	0	0	0	0	0	2	24	0	0	0	40	2	0	0	211	0
3	0	0	32	554	104	3	0	0	0	0	0	0	0	0	92	0	1	0	41	0	0	0	436	0
4	0	0	0	311	758	6	141	0	0	0	0	0	0	2	175	0	1	0	30	58	1	0	157	0
5	0	0	0	143	0	0	0	74	0	0	0	0	0	0	106	0	0	0	112	0	0	0	0	0
6	0	0	0	186	0	0	0	135	0	0	0	0	0	0	103	0	0	6	174	0	0	1	0	0
7	0	0	0	3	416	3	0	0	0	0	0	0	0	0	52	0	18	4	2	4	0	0	70	0
8	0	0	0	32	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	19	4	19	0	0	0	0	0	0	0	0	2	44	0	0	0	0	0	0	14	0
10	0	0	0	0	34	27	0	0	0	0	0	0	0	0	1	0	0	0	19	1	0	0	2	0
11	0	0	0	14	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
12	0	0	0	134	12	0	0	0	0	0	1	0	0	0	0	18	0	0	32	0	0	0	3	0
13	0	0	0	35	89	0	5	0	0	0	0	0	0	0	0	43	0	4	14	0	0	0	84	0
14	0	0	0	145	136	0	0	0	0	0	0	0	0	0	3	67	2	6	0	7	0	0	25	0
15	0	0	0	15	146	0	0	0	0	0	0	0	0	0	1	4	21	0	0	1	0	0	30	0
16	0	0	0	484	1	0	0	79	0	0	0	0	0	0	0	14	0	0	108	0	0	0	0	0
17	0	0	0	0	146	5	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	71	0
18	0	0	0	127	311	0	15	0	0	0	0	0	0	0	2	108	1	21	5	71	19	0	0	82
19	0	0	0	0	134	16	0	0	0	0	0	0	0	0	0	59	0	1	0	38	0	0	54	0
20	0	0	0	49	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0
21	0	0	0	486	179	90	16	0	0	0	0	0	0	0	0	16	19	28	6	43	2	0	0	187
22	0	0	0	0	11	14	3	40	855	0	0	0	0	0	0	68	0	21	0	132	0	0	7	1
23	0	0	0	80	0	0	0	0	20	0	0	0	0	0	0	53	0	0	0	20	0	0	0	0
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	



Μελλοντικές Κατευθύνσεις



Πειραματισμός με νέες
αρχιτεκτονικές μοντέλων.



Αντικατάσταση του
oversampling με data
augmentation.



Ανάπτυξη μοντέλου object
detection σε πραγματικό
χρόνο.

Βιβλιογραφία

- [1] Aganian, D. *et al.* (2023) *How object information improves skeleton-based human action recognition in assembly tasks*, *arXiv.org*. Available at: <https://arxiv.org/abs/2306.05844> (Accessed: 30 September 2024).
- [2] (No date a) *A systematic literature review on vision based gesture recognition techniques | request PDF*. Available at: https://www.researchgate.net/publication/324805847_A_systematic_literature_review_on_vision_based_gesture_recognition_techniques (Accessed: 30 September 2024).
- [3] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020, June 17). *BlazePose: On-device Real-time Body Pose tracking*. *arXiv.org*. <https://arxiv.org/abs/2006.10204>
- [4] Author links open overlay panelRoshan Singh a *et al.* (2022) *Recent trends in Human activity recognition – a comparative study*, *Cognitive Systems Research*. Available at: <https://www.sciencedirect.com/science/article/pii/S138904172200047X> (Accessed: 30 September 2024).
- [5] Ben-Shabat, Y. *et al.* (2023) *The IKEA ASM dataset: Understanding people assembling furniture through actions, objects and pose*, *arXiv.org*. Available at: <https://arxiv.org/abs/2007.00394> (Accessed: 30 September 2024).
- [6] (No date) *(PDF) real-time human action recognition using Deep Learning*. Available at: https://www.researchgate.net/publication/366523828_Real-Time_Human_Action_Recognition_Using_Deep_Learning (Accessed: 30 September 2024).