

---

## INDICE

---

1	L'apprendimento automatico	11
1.1	Deep learning	13
1.2	Le reti neurali	13
1.3	Le reti neurali ricorrenti	16
1.4	Le reti LSTM	18
2	Lavori precedenti	23
2.1	Recognizing Human Actions as Evolution of Pose Estimation Maps (2018)	24
2.2	Action Machine: Rethinking Action Recognition in Trimmed Videos (2018)	28
2.3	Overview	31
3	Stima della posa	33
3.1	PoseNet	34
3.1.1	Stima dei keypoint	36
3.1.2	Raggruppamento dei keypoint in istanze di persona	37
3.2	Detectron-2	38
3.2.1	Mask R-CNN	41
3.2.2	Predizione della posa con Mask R-CNN	42
4	Metodo proposto	45
4.1	Estrazione delle pose	45
4.2	Assegnazione coerente delle pose	46
4.3	Rimozione degli zeri	48
4.4	Tecniche di rielaborazione	49
4.4.1	Baricentro del frame	50
4.4.2	Baricentro del video	51
4.4.3	Baricentri multipli	52
4.4.4	Tecnica "Next frame"	54
4.4.5	Tecnica delle distanze cumulate	55
4.4.6	Tecnica delle distanze relative	56
4.5	Normalizzazione	57
4.6	Struttura della rete	58
4.7	Criterio di arresto dell'addestramento	60
5	Il dataset NTU RGB+D	63
5.1	NTU RGB+D in dettaglio	65

## 2 Indice

5.2	Criteri di valutazione	67
5.2.1	Valutazione Cross-subject	67
5.2.2	Valutazione Cross-view	68
6	Esperimenti e risultati	69
6.1	Una porzione rappresentativa del dataset	70
6.2	Fase 1: Test di tutte le tecniche	70
6.3	Fase 2: Numero di livelli	72
6.4	Fase 3: Dropout e regolarizzatori	76
6.5	Fase 4: Allenamento completo tecnica migliore	76
7	Conclusioni e sviluppi futuri	77
7.1	Sviluppi futuri	79