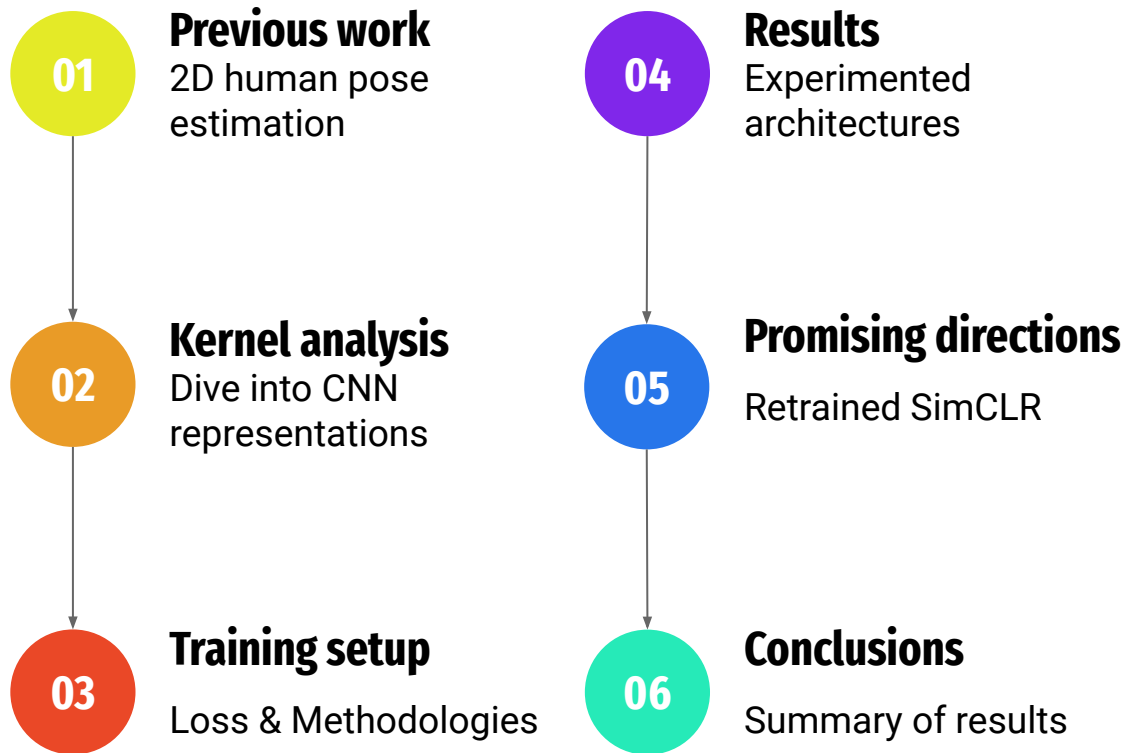


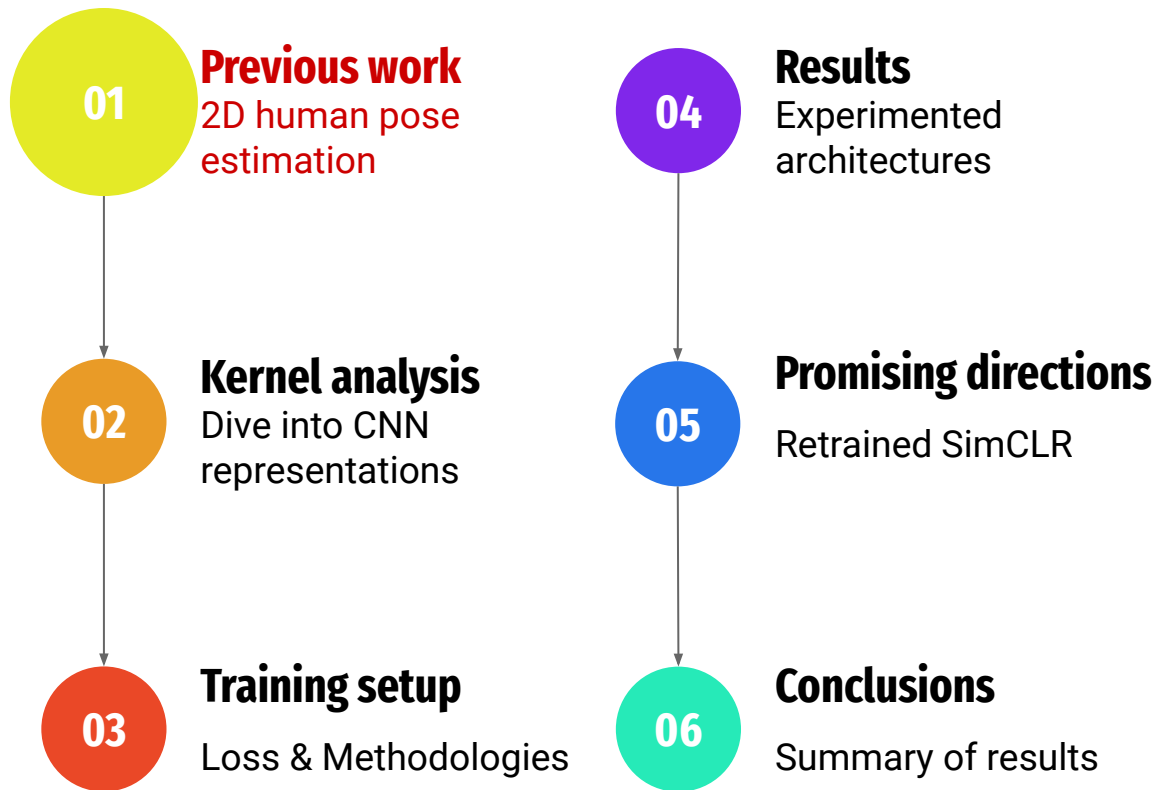
Human pose estimation using contrastive learning

SimCLR & SimSiam
for 3D Human pose estimation

Summary

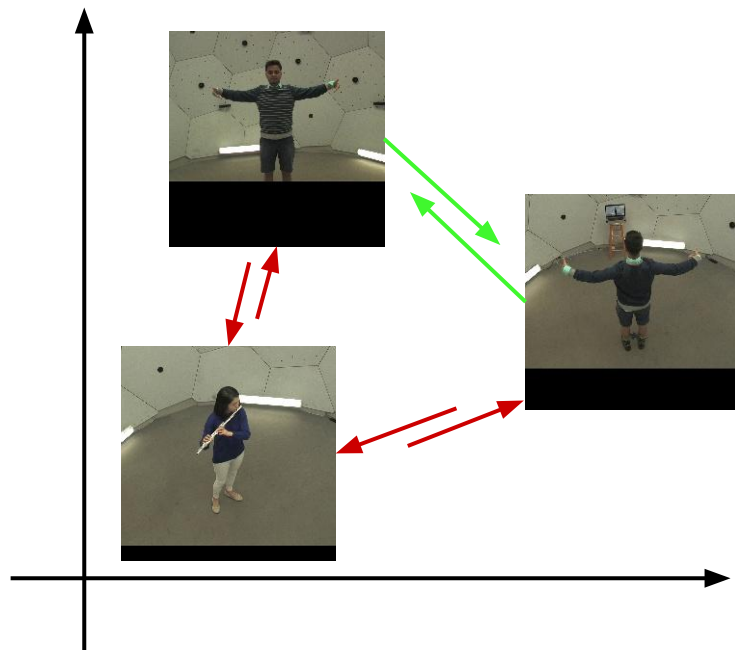


Summary



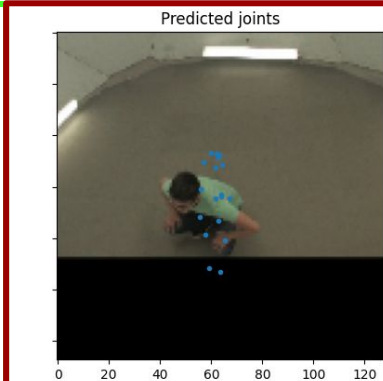
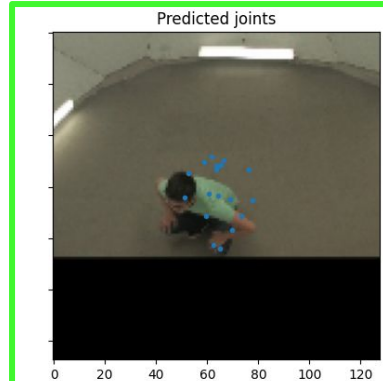
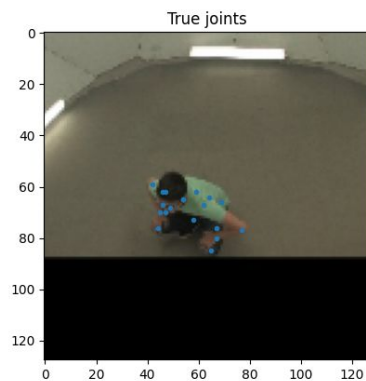
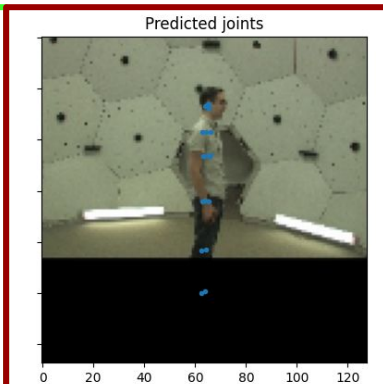
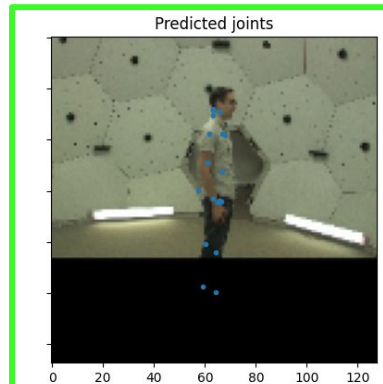
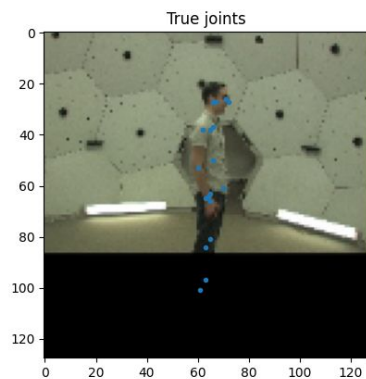
SimCLR & SimSiam

	SimCLR	SimSiam
Loss	Positive Examples + Negative Examples	Positive Examples
Model architecture	CNN base encoder (ResNet50) + projection head	CNN base encoder (ResNet50) + projection head + prediction head
Training details	20 epochs with batch size 200 0.01 learning rate for projection head 0.01/2 learning rate for base encoder	20 epochs with batch size 256 0.05 learning rate for all layers
Silhouette score	0.45 for encoder features 0.53 for projection head	0.48 for encoder features 0.53 for projection head 0.53 for prediction head

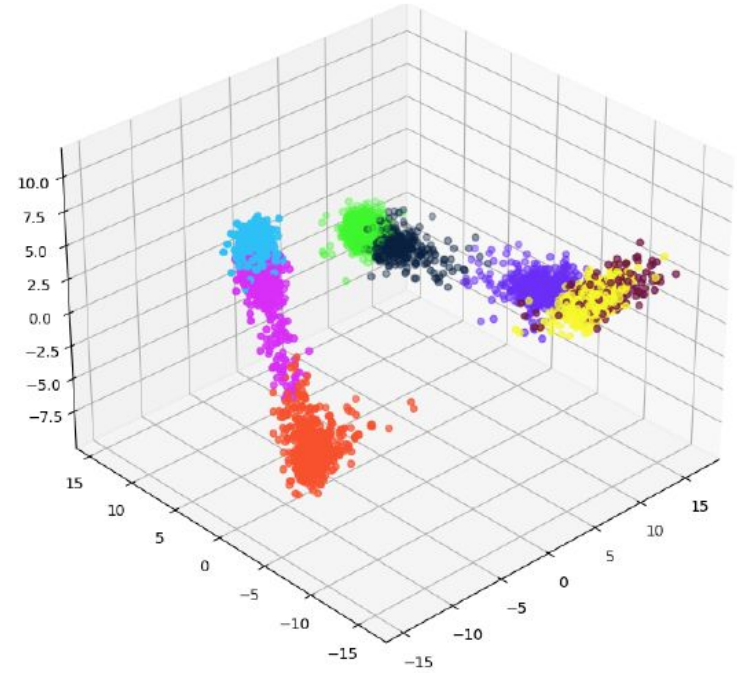
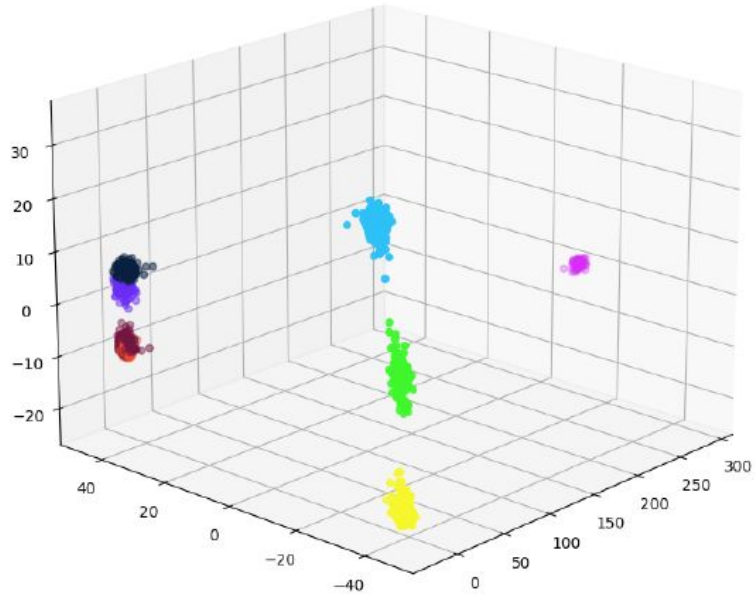


Results 2D regression

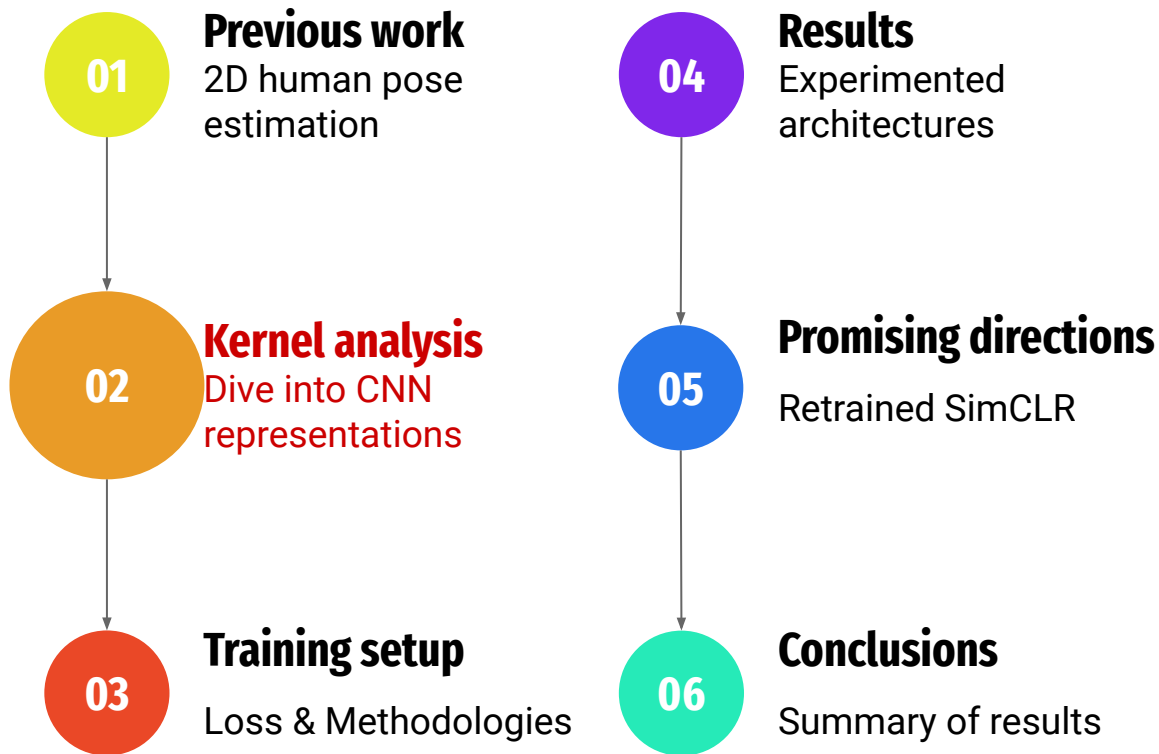
Model	Training	Validation	Test
SimCLR	0.4431	0.4442	0.4440
SimSiam	0.4075	0.4098	0.4111



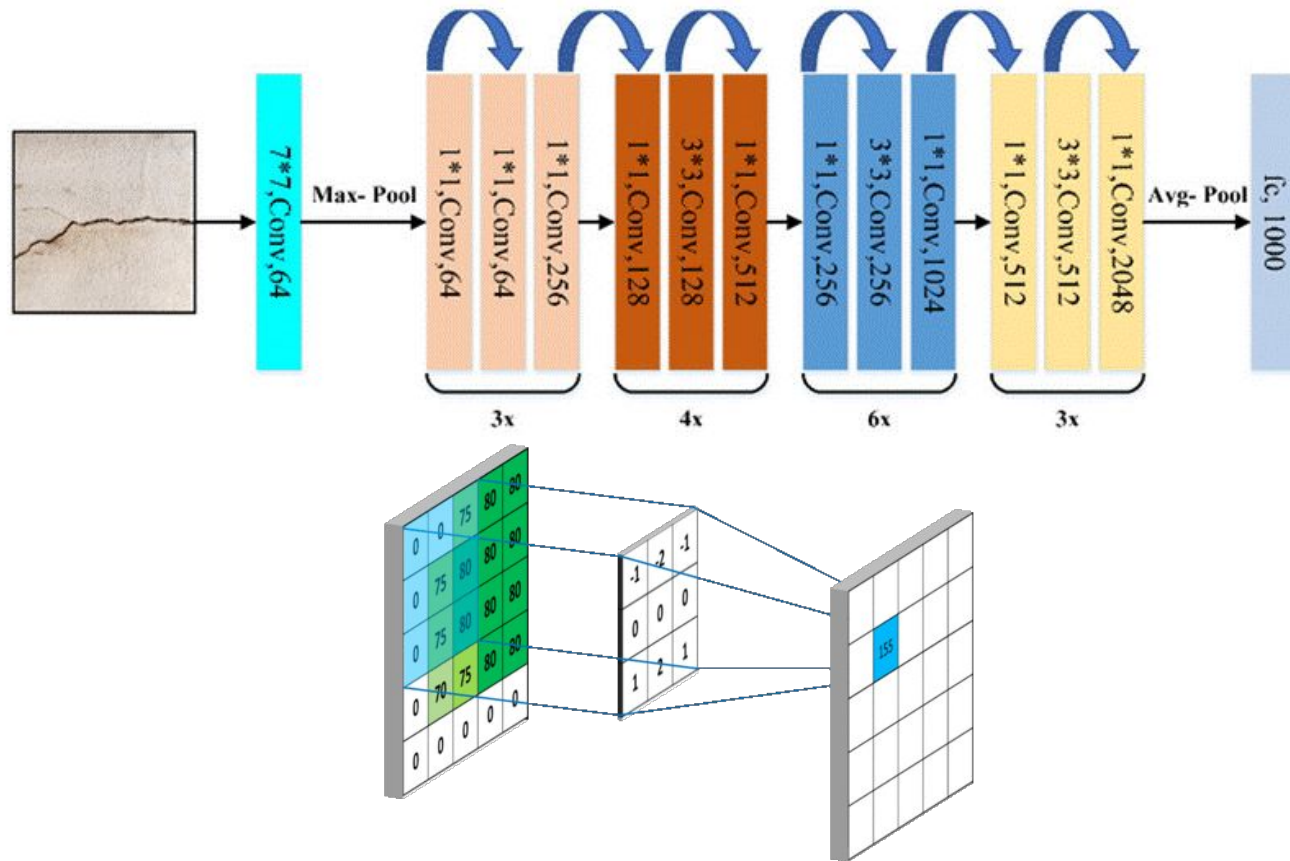
Criticality



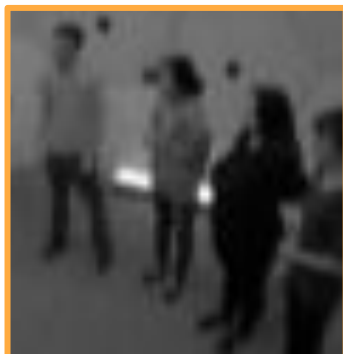
Summary



ResNet50 architecture and CNN

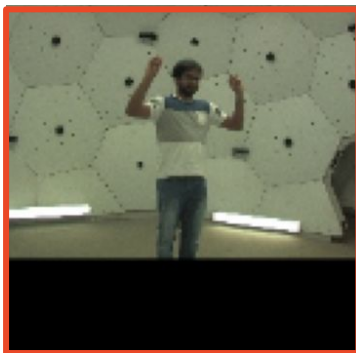


LPF



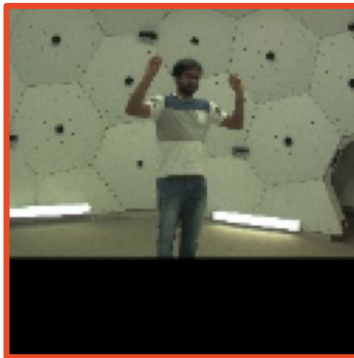
$$\begin{bmatrix} 0.0023 & -0.0321 & -0.0039 & -0.0125 & -0.0851 & -0.0224 & 0.1448 \\ 0.0298 & 0.0104 & -0.0041 & -0.0536 & -0.3131 & -0.3287 & 0.0601 \\ 0.0111 & 0.0092 & 0.1226 & 0.1340 & -0.3704 & -0.5951 & -0.2644 \\ 0.0573 & 0.0649 & 0.3319 & 0.6767 & 0.2513 & -0.2713 & -0.2088 \\ -0.0967 & -0.1151 & 0.1277 & 0.4881 & 0.4082 & -0.0066 & -0.0822 \\ -0.0856 & -0.2264 & -0.0531 & 0.2177 & 0.2357 & 0.0021 & 0.0172 \\ 0.0532 & -0.1861 & -0.1513 & 0.0275 & 0.1330 & -0.0254 & -0.0156 \end{bmatrix}$$

HPF

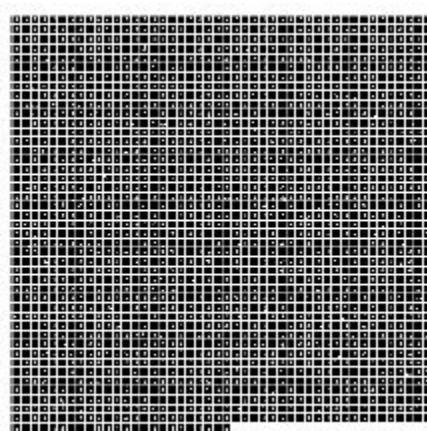
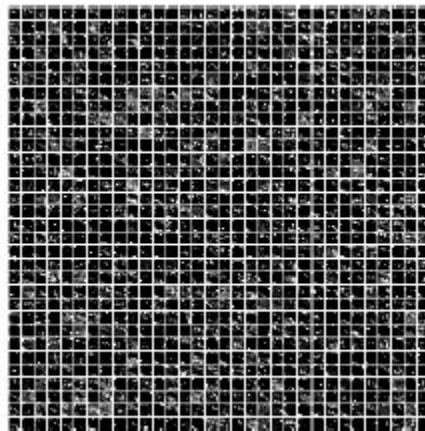
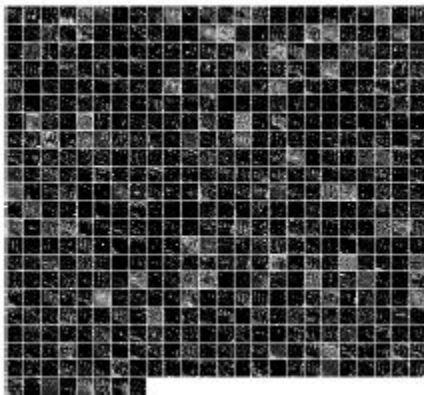
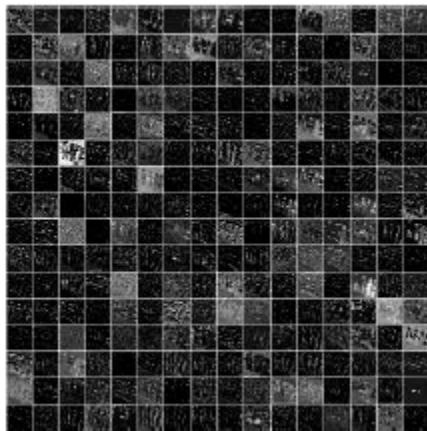


$$\begin{bmatrix} 0.0269 & -0.0692 & -0.0056 & 0.0635 & -0.0635 & -0.0081 & 0.0346 \\ 0.0040 & -0.0071 & 0.1616 & 0.2279 & 0.0785 & 0.0396 & -0.0044 \\ -0.0517 & 0.0185 & 0.2755 & 0.3411 & 0.1082 & 0.0609 & 0.0029 \\ -0.0407 & 0.0345 & 0.2471 & 0.2760 & 0.0752 & 0.0241 & 0.0355 \\ -0.0399 & -0.0210 & 0.1453 & 0.1981 & 0.0657 & -0.0254 & -0.0307 \\ 0.0087 & -0.0295 & 0.0759 & 0.1380 & 0.0080 & -0.0373 & -0.0433 \\ 0.0190 & -0.0311 & 0.0251 & 0.1204 & 0.0474 & -0.0036 & -0.0005 \end{bmatrix}$$

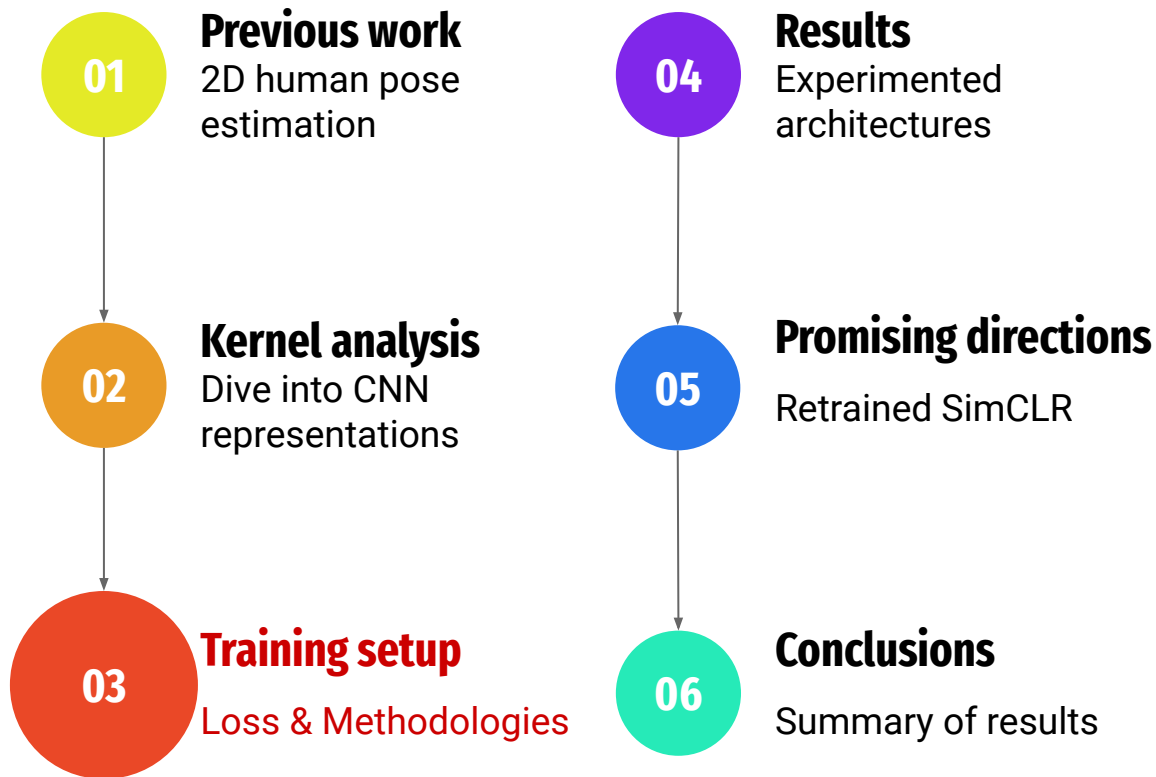
Negative



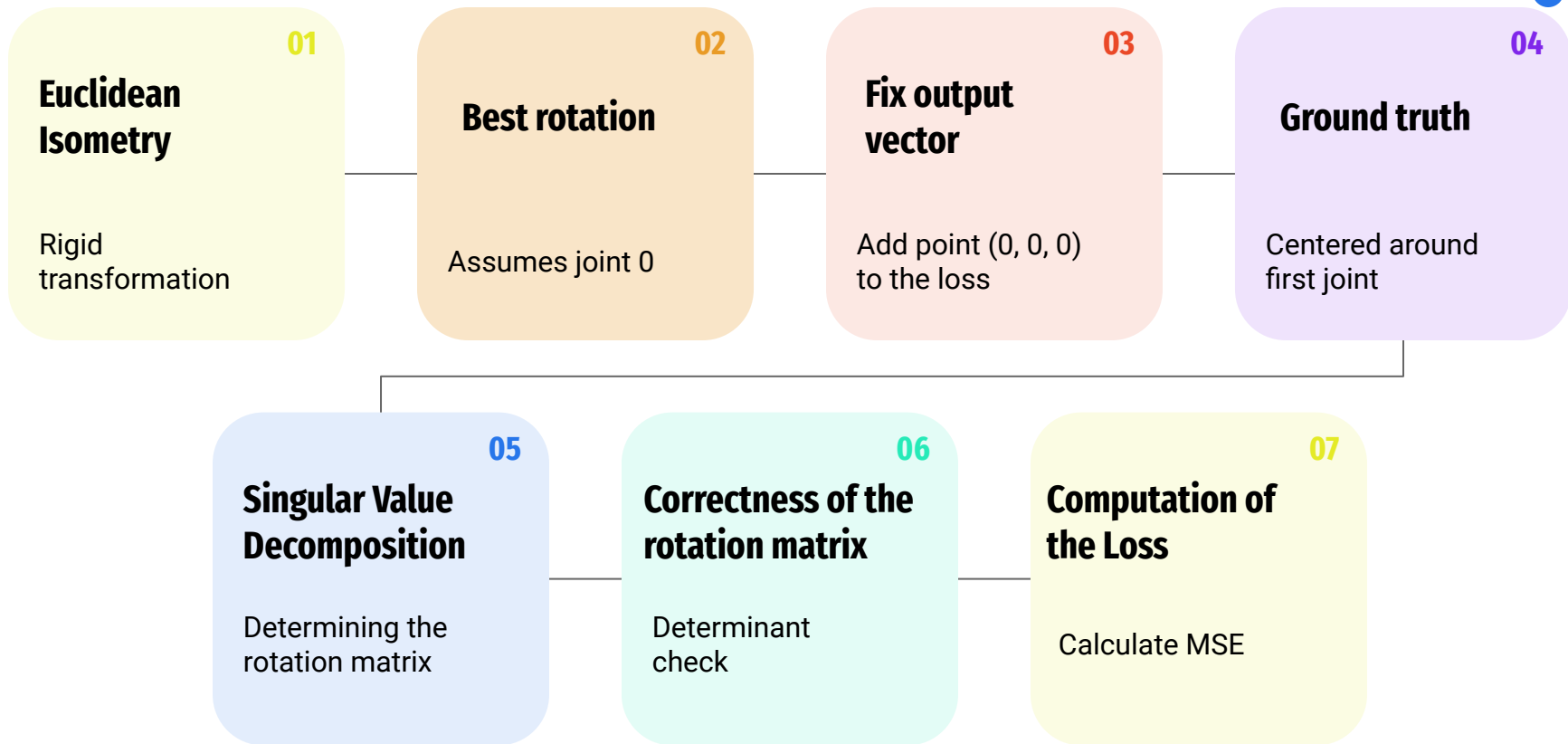
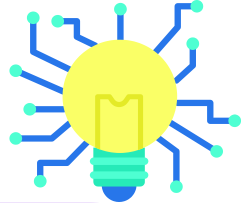
Inner layers



Summary



Loss



Training setup



A diagram illustrating the training setup. A single point on the left branches out into four grey triangular shapes, each pointing to a specific parameter. The parameters are listed in colored rounded rectangles, each with a corresponding detail in a larger, lighter-colored rounded rectangle to its right.

Cosine annealing

LR scheduler

Epochs

We opted for 30

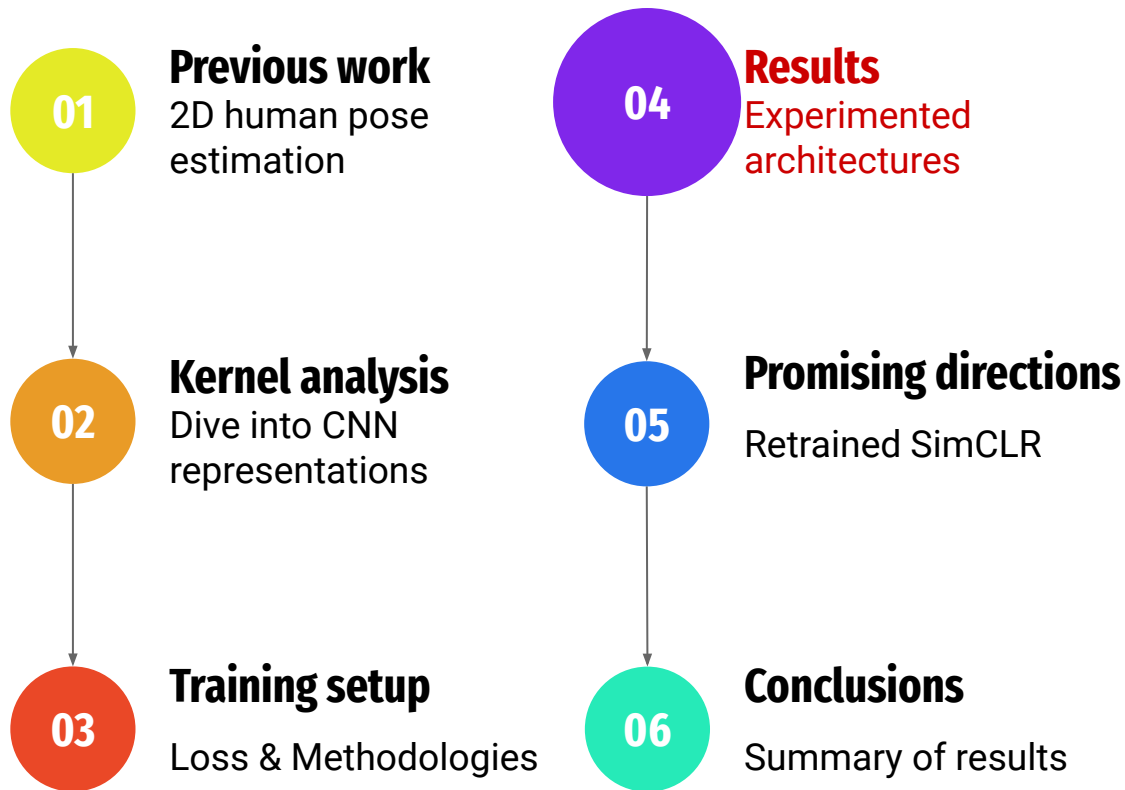
Initial Learning Rate

Smaller LR of 0.02

Batch size

Same batch size of 128

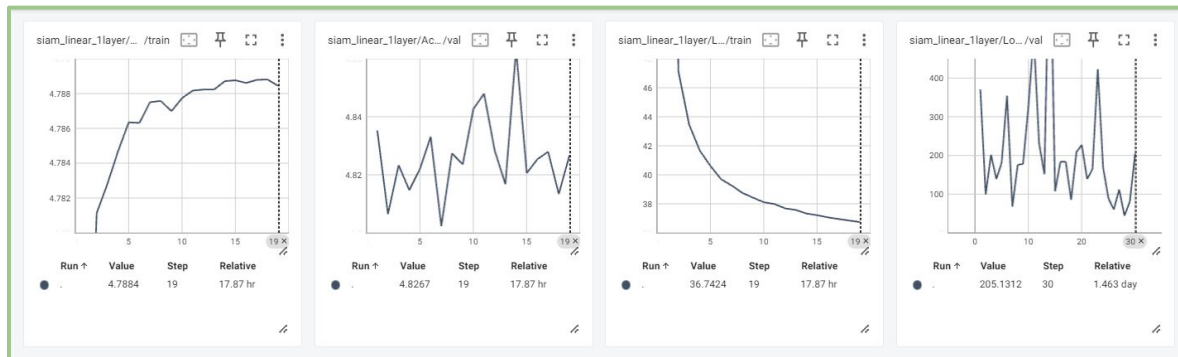
Summary



1-layer architecture & results

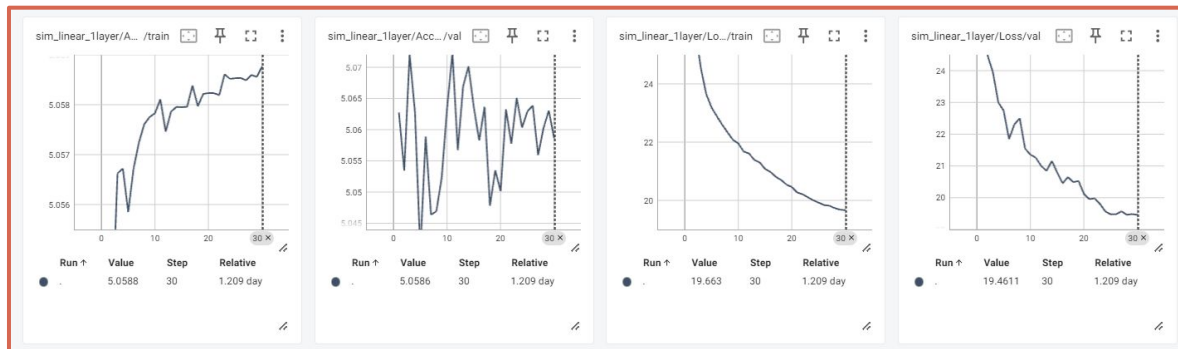
01 SimSiam

	training	validation	testing
Loss	243.46	205.13	233.51
Accuracy	4.82	4.82	4.82



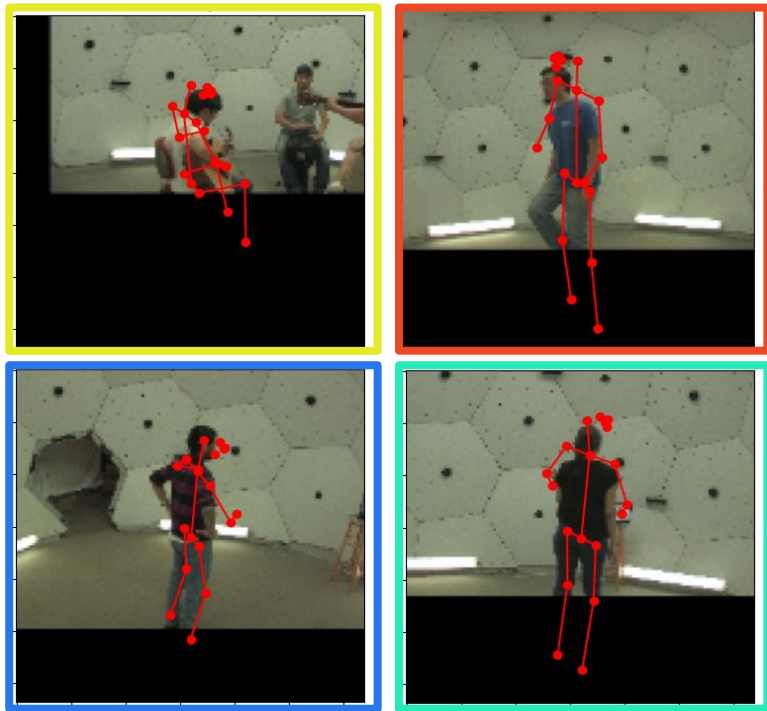
02 SimCLR

	training	validation	testing
Loss	19.14	19.46	19.64
Accuracy	5.05	5.96	5.06

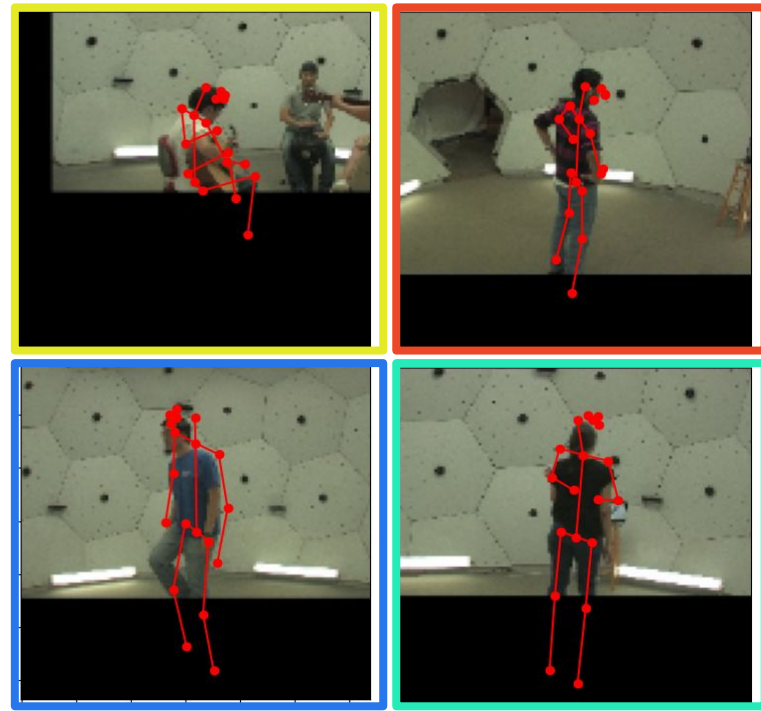


Qualitative analysis

01 SimSiam



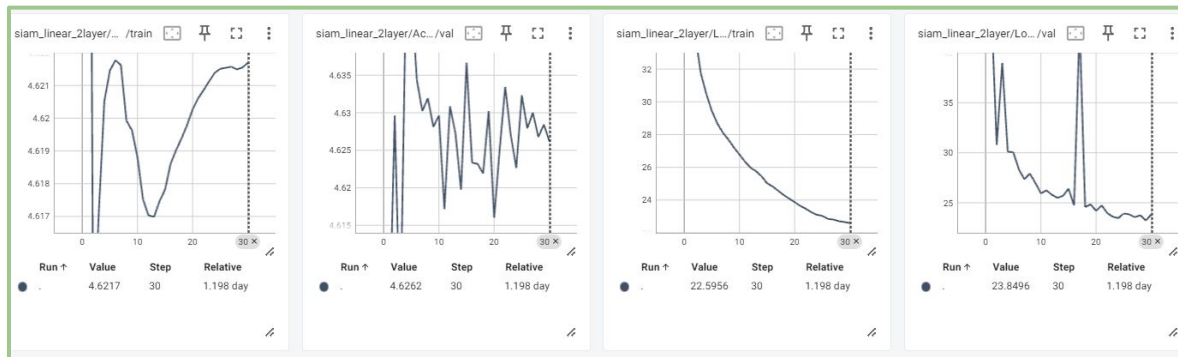
02 SimCLR



2-layer architecture & results

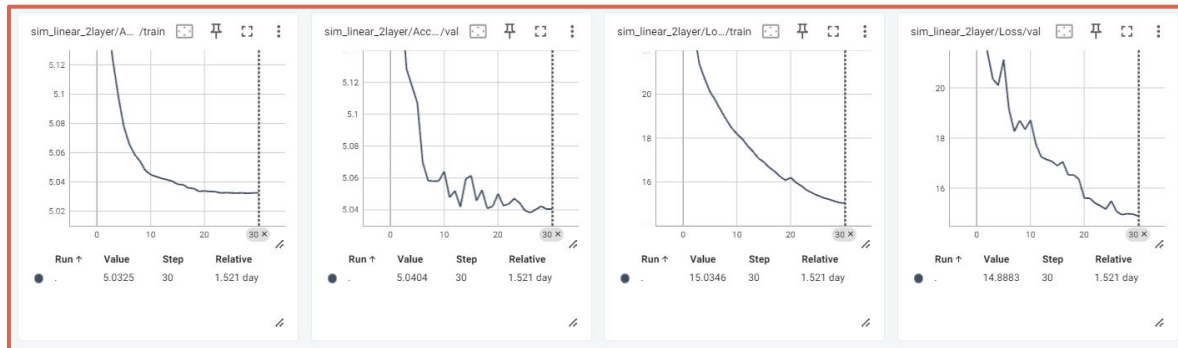
01 SimSiam

	training	validation	testing
Loss	22.32	23.85	24.22
Accuracy	4.62	4.63	4.63



02 SimCLR

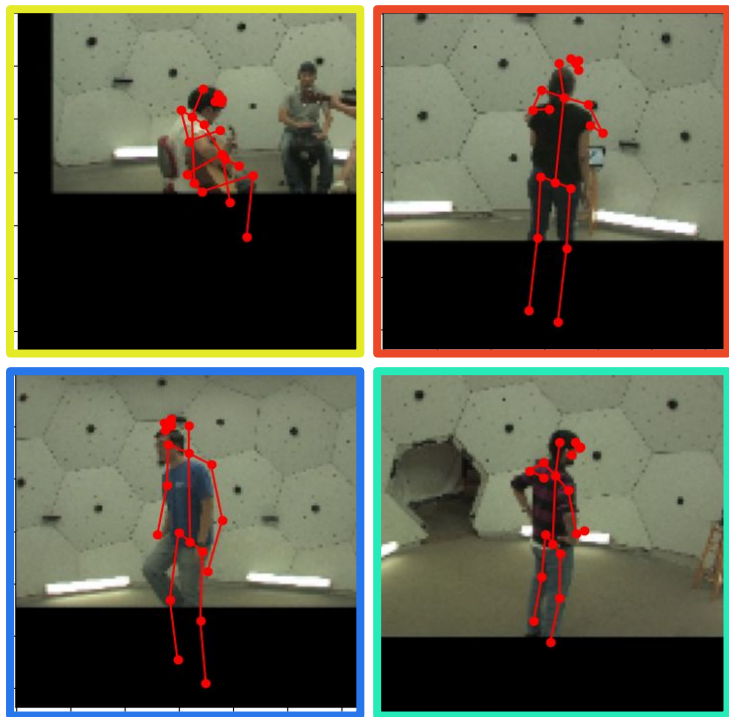
	training	validation	testing
Loss	16.61	14.89	14.96
Accuracy	5.03	5.04	5.04



Qualitative analysis

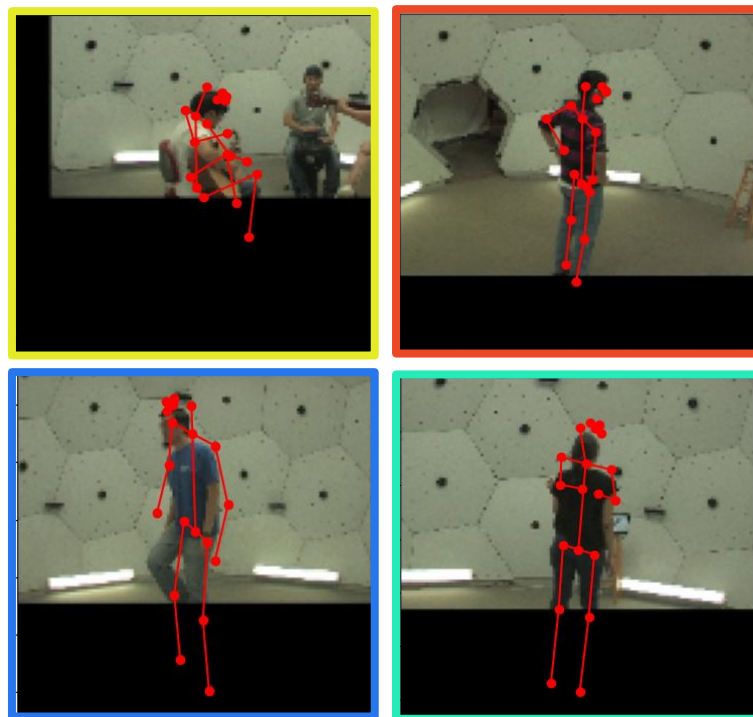
01

SimSiam

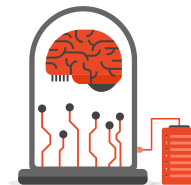


02

SimCLR

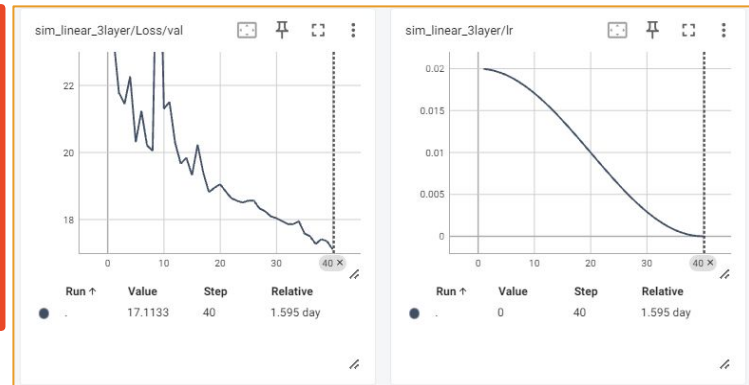


3-layer architecture & results

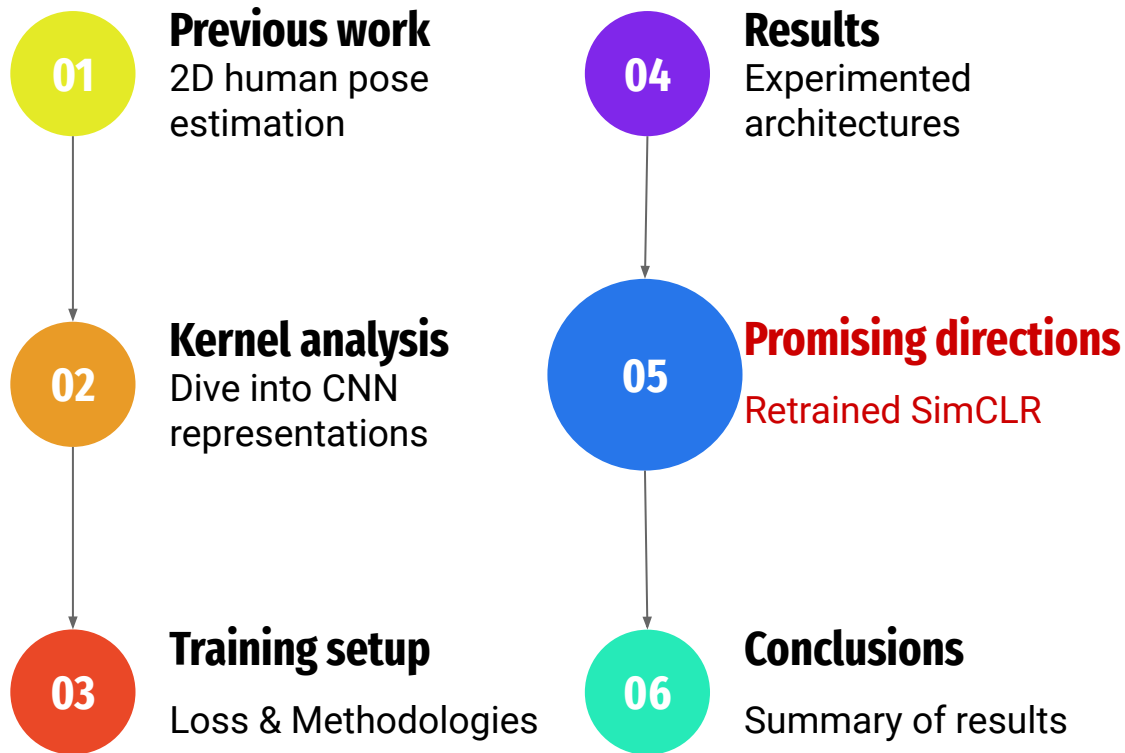


Result SimCLR

	training	validation	testing
Loss	16.62	17.11	17.20
Accuracy	5.11	5.12	5.12



Summary



Training setup

01 22 Epochs

After 22 epochs the loss remained constant, maybe should have used bigger eta_min

02 225 Batch size

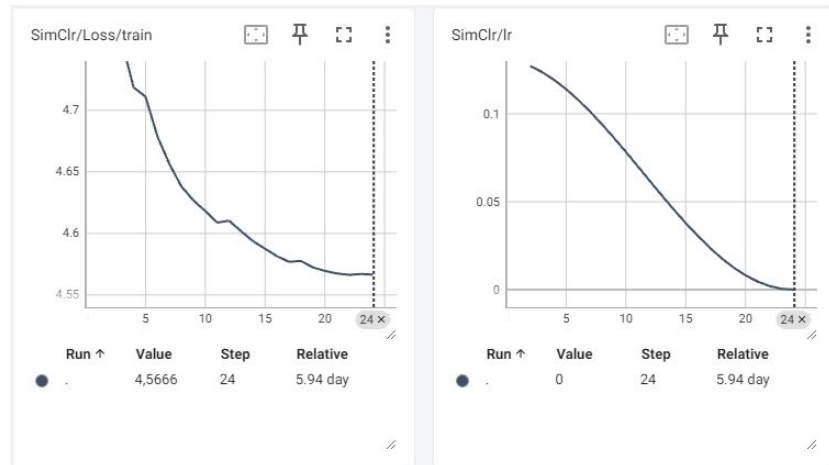
Batch size similar to previous, still problem of negative pairs

03 0.3×batch size/256 initial LR

Bigger initial LR

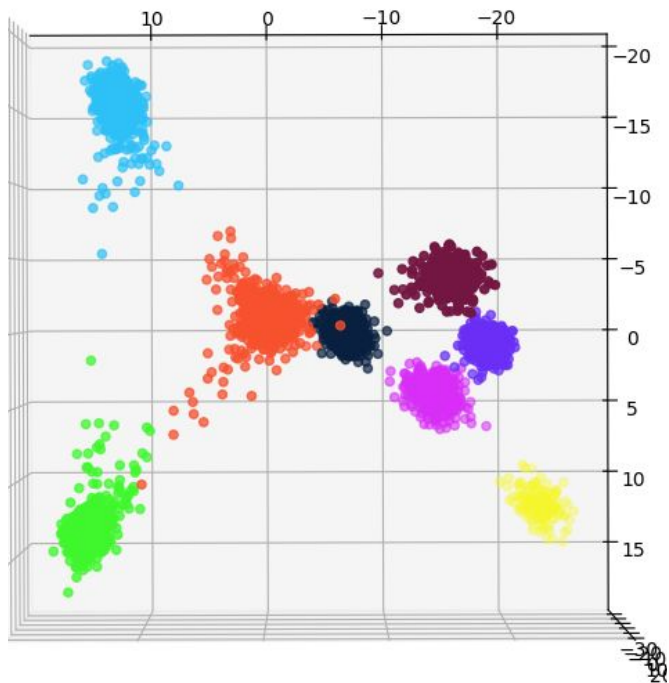
04 Cosine Annealing T_max=25

LR scheduling

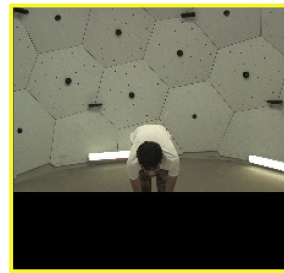
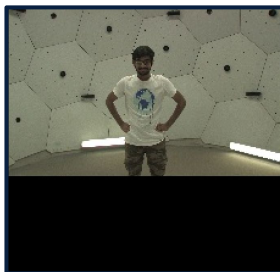
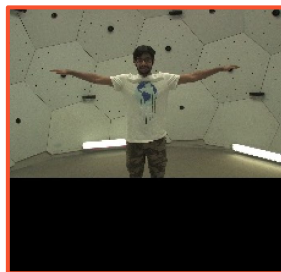
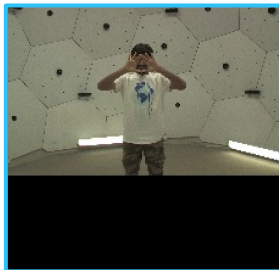


	silhouette score
encoder features	0.30
projector features	0.48

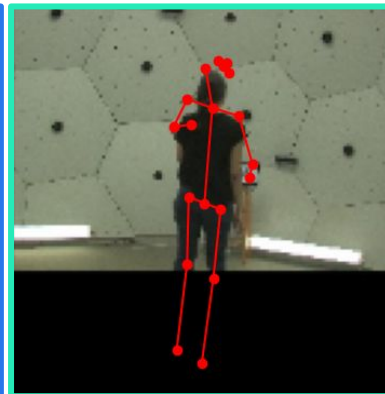
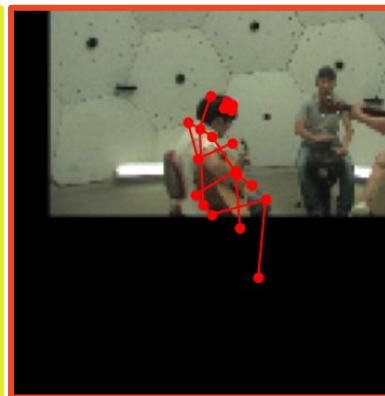
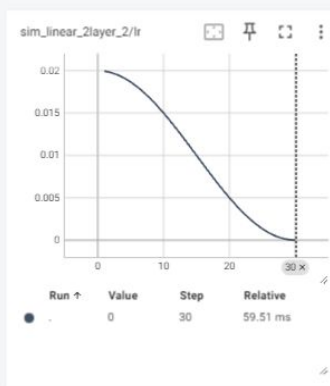
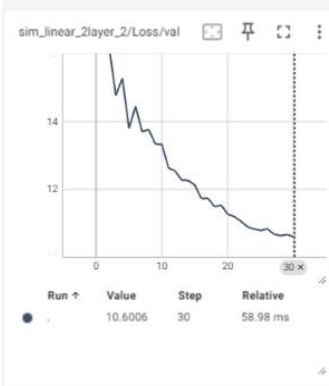
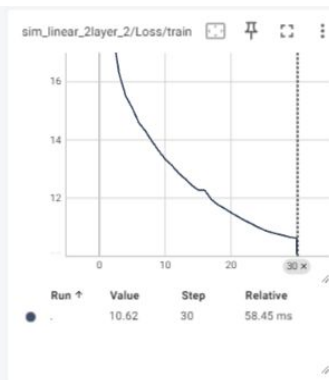
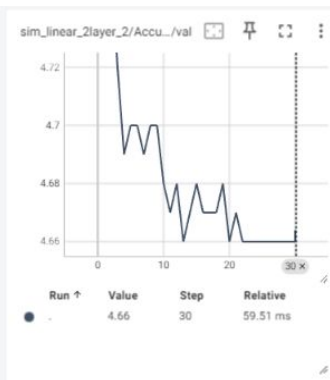
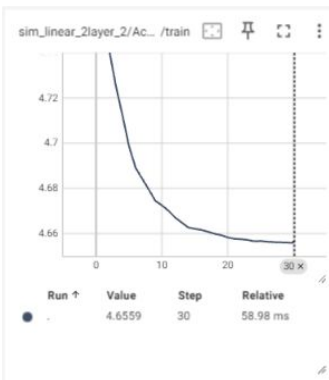
Encoding Space



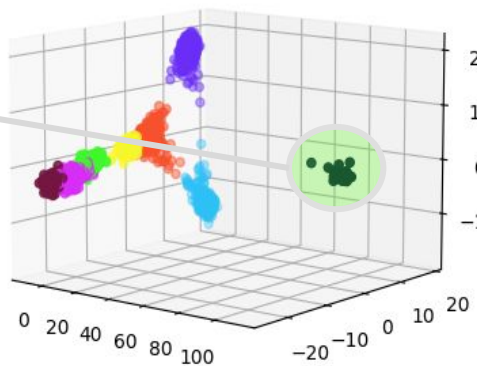
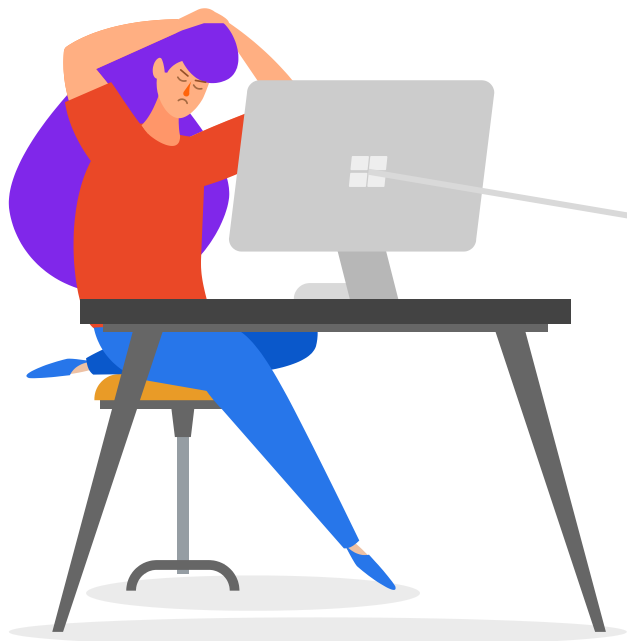
(obtained by K means clustering with random_state=3)



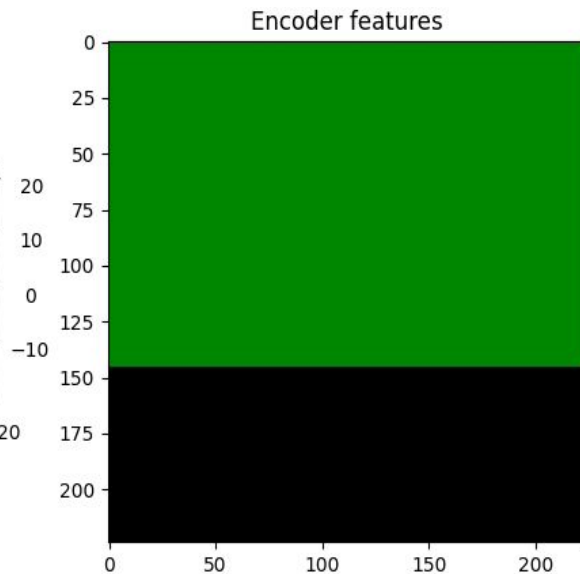
Results



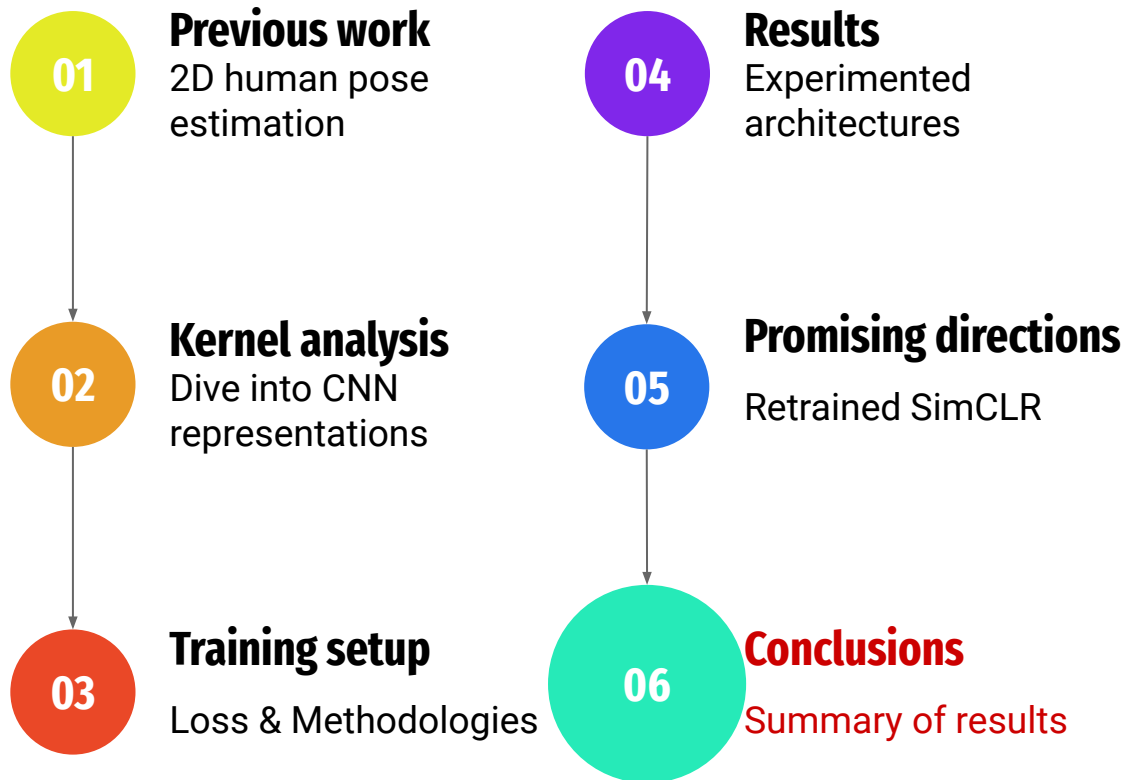
Known Issues



(obtained by K means clustering with random_state=6)



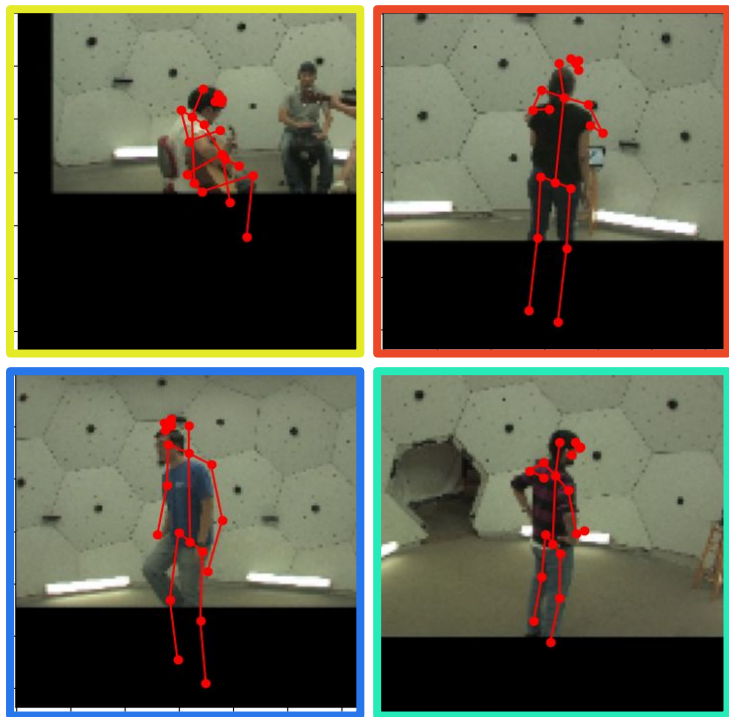
Summary



Contrastive Learning for Human Pose Estimation

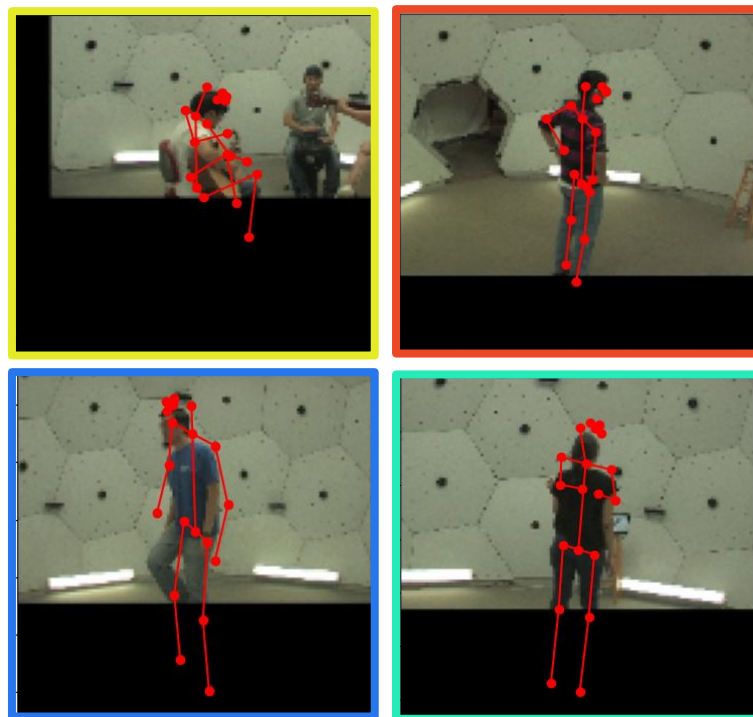
01

SimSiam



02

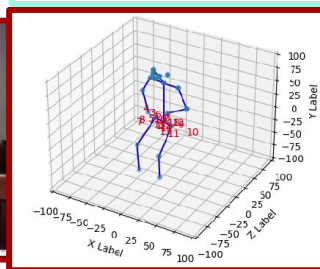
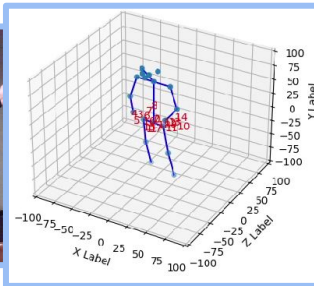
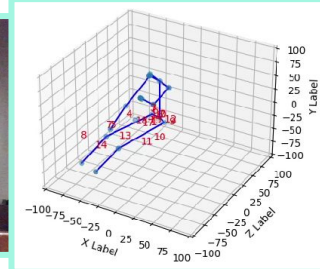
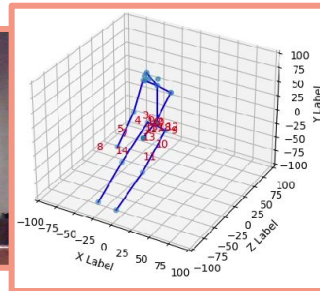
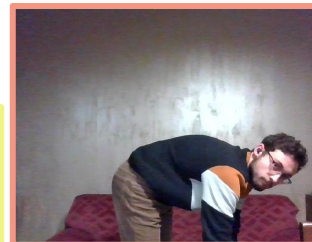
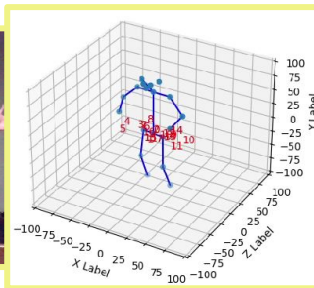
SimCLR



Future work & Improvements

Worse with camera feed

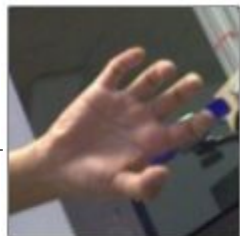
Less capable with unseen scenes and uncentered subject



Future work & Improvements

01 Negative pairs

Bigger batch size allowing for more negative pairs examples



Original



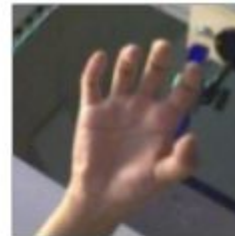
Color jitter



Cut out



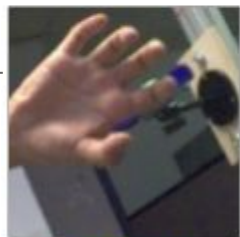
Scale



Rotate

02 Data augmentation

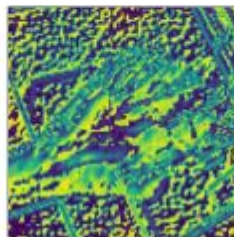
Use some transformations on images to increase generalization capability



Translate



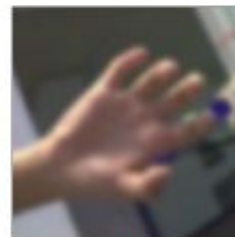
Noise



Sobel filter



Color drop



Blur

[3D hand pose estimation](#)

THANK YOU

for your attention

Alessia Pivotto
Davide Cavicchini
Sofia Lorengo

