

# Master Thesis Questions

André Oskar Andersen

1. Hvorfor bruger man heatmaps?

Heatmaps bruges til at modellere usikkerheden der er ved annoteringen

2. Hvorfor bruger PCK en threshold baseret på torso height?

Det gør man for at inkorporere størrelsen af subjektet. Havde man fikset thresholden, havde predictionen været meget nemmere når personen er lille i billedet.

3. Hvorfor kigger man på om afstanden ligger under en radius?

Ligesom ved årsagen til brugen af heatmaps, så modellerer thresholden usikkerheden der er ved annoteringen.

4. Hvorfor gør du brug af MSE?

DeciWatch gør brug af en modificeret udgave af MSE. Unipose-LSTM gør brug af MSE. For at holde det konstant, valgte jeg bare at bruge MSE for alle modellerne.

5. Kan du forklare ADAM?

6. Hvad er fordelen ved at bruge ReLU?

Vigtigst af alt, lader det modellerne modellere ikke-lineære funktioner. Yderligere, så er der nogle andre fordele, såsom (1) ReLU er hurtigt at udregne, (2) den afledte er nemt at udregne og (3) modsat andre activation functions såsom Sigmoid, har den en mindre sandsynlighed for at introducere vanishing gradients

7. Hvorfor hjælper layer normalization

Ved at normalisere lagene af dataen undgår vi at inputtet ikke "ping-ponger" frem og tilbage, så dataen ikke pludseligt forskydes, idet den centrerer dataen. Dette burde sørge for, at færre epochs skal bruges.

8. Hvorfor gøres der brug af 1x1 convolutions

1x1 convolutions bruges som et fully-connected layer på tværs af filterne. Det bruges ofte til at downsample antallet af filtre.

9. Hvorfor initialiserer med Glorot?

Der er flere grunde til det. (1) vægtene hverken for store eller små, hvilket hjælper på vanishing/exploding gradients. (2) vi sørger for at der ikke er nogen symmetri, hvilket ville resultere i neuronerne have samme udregninger.

10. Hvilke andre metoder kunne man bruge?

11. Spørgsmål om Convolution, LSTM og Transformer

12. Du siger at du har introduceret noget evaluation bias i pretraining, da forskellige frames fra den samme video sequence kan optræde i forskellige subsets og den samme person derved optræder på tværs af subsets. Har du ikke samme problem i finetuning?

Både jo og nej. Jo, fordi den samme person kan optræde på tværs af subsets. Nej, fordi det er forskellige video sequences.

13. Hvorfor er 3DConv den bedst performing model?

14. Hvordan performer din deciwatch i forholdet til det fra artiklen?

15. Hvilke andre metoder findes der til at undgå overfitting?

16. Hvorfor brugte du ikke data augmentation?

17. Hvorfor bruger DeciWatch ikke alle frames?

18. Hvorfor er fem frames det optimale valg?