1. **What is a "skip connection"?**

værdierne af et output bliver til inputtet på flere lag, hvor der ikke bliver gjort noget ved disse, lag bliver altså “skippet”, når en gradient kommer for tæt på 0.

1. **Why do skip connections allow us to train deeper models?**

Det gør det muligt at undgå gradient degradation, således at der ikke kan være for mange lag i en model.

1. **What is "identity mapping"?**

Returnerer inputtet uden at ændre det

1. **What do ResNets have to do with residuals?**

ResNets prøver ikke at komme med et output fra et layer, men i stedet for med at minimere differencen mellem outputtet og det ønskede resultat. Dermed er ResNets god til at lære om der er små ændringer eller om der skal skippes.

“*If the outcome of a given layer is x and we’re using a ResNet block that returns y = x + block(x), we’re not asking the block to predict y; we are asking it to predict the difference between y and x. So the job of those blocks isn’t to predict certain fea‐ tures, but to minimize the error between x and the desired y. A ResNet is, therefore, good at learning about slight differences between doing nothing and passing through a block of two convolutional layers (with trainable weights). This is how these models got their name: they’re predicting residuals (reminder: “residual” is prediction minus target).*” - fra side 448

1. **How do we deal with the skip connection when there is a stride-2 convolution? How about when the number of filters changes?**

Ved at bruge *fully convolutional networks* hvor man tager gennemsnittet af activations henover et convolutional grid, dermed bliver et grid af flere activation til én activation pr. billede.

1. **What is the "stem" of a CNN?**

Stammen er starten af et CNN som oftest har en anden struktur end resten.

Grunden til det er at man fandt ud af at størstedelen af beregning sker i de første lag. Derfor prøver man at holde stammen så simpelt og hurtigt som muligt

1. **How does a bottleneck block differ from a plain ResNet block?**

Bottleneck layers bruger 3 forskellige convolutions 2 1x1 i begyndelsen og slutningen, og en 3x3, istedet for 2 3x3.

1. **Why is a bottleneck block faster?**

Bottleneck blocks reducerer antallet af parametre og matrix multiplications.

Hvis de 50 minutter ikke er gået så besvar følgende:

1. What is "adaptive pooling"?
2. What is "average pooling"?
3. Why do we need Flatten after an adaptive average pooling layer?
4. Why do we use plain convolutions in the CNN stem, instead of ResNet blocks?
5. How do fully convolutional nets (and nets with adaptive pooling in general) allow for progressive resizing?