Transfer Learning
Aside: Skip Connections

```
Aside: Skip Connections

Conv
ReLU
+
Conv
ReLU
cat
Conv
Conv
Conv
UpConv
UpConv
UpConv
3, 32, 32
8, 16, 16
16, 8, 8
32, 4, 4
32, 4, 4
16, 8, 8
32, 8, 8
16, 8, 8
8, 16, 16
16, 16, 16
1, 32, 32
```
Transfer Learning

Small dataset:
- Conv
- ReLU
- Conv
- ReLU
- Linear

Overfits

Large dataset:
- Conv
- ReLU
- Conv
- ReLU
- Linear

Pre-training

Different, related task

Copy weights

Overfits less

Original dataset:
- Conv
- ReLU
- Conv
- ReLU
- Linear

Train just a little

Fine tuning
Why does Transfer Learning Work?

Similar datasets – similar features

- Edges
- Colors
- Basic shapes
- Textures
- ...

- Edges
- Colors
- Basic shapes
- Textures
- ...
Pre-training

- Vision: ImageNet
- Natural Language Processing
  - Self-Supervision
  - GPT3
- Model zoo
  - torch.utils.model_zoo
Fine Tuning

Large dataset:

Original dataset:

Copy weights

Conv
ReLU
Conv
ReLU
Linear

Conv
ReLU
Conv
ReLU
Linear

Sometimes reinitialize the last few layers
When to use Transfer Learning

- Whenever you can
  - If a pre-trained model exists
  - Especially in early experiments
- In this class: don’t use pretrained models
  - You can look at them for early experiments if you want
  - Your final models should be trained from scratch