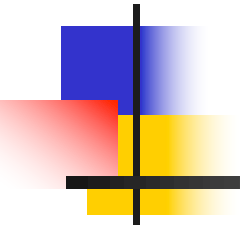
CMSC 510 – L20
Regularization Methods for Machine Learning



Part 20b: Skip connections

Instructor:
Dr. Tom Arodz

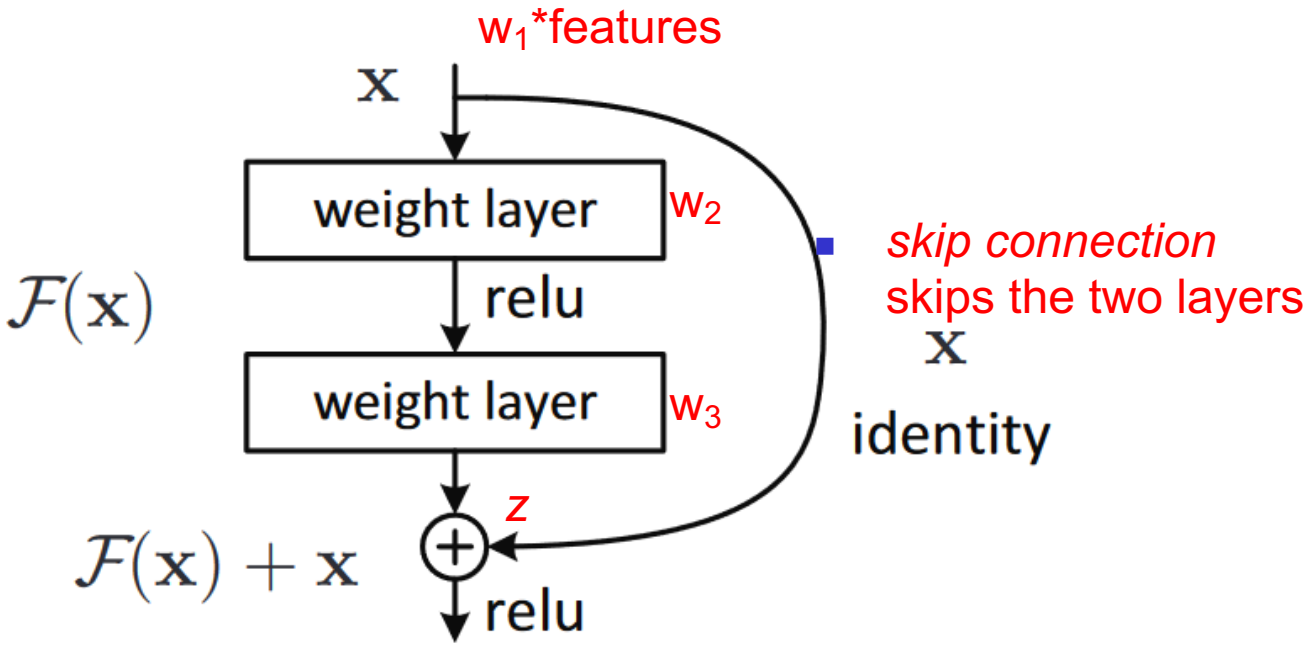
Multi-layer networks

- A neural network with ReLU activation is essentially:
  - $Y=\mathrm{ReLU}(\ W_3\,\mathrm{ReLU}(\ W_2\,\mathrm{ReLU}(\ W_1X\ )\ )\ )$
- We see terms like $z=w_{3ij}*w_{2kl}*w_{1mn}*x$
- The derivative of z over w_{2kl} is $w_{3ij}*\ w_{1mn}\ *x$
- If w_{3ij} is small
  the gradient "signal" telling w_{2kl} in which way to change does not reach w_{2kl}

Skip connections

- $z = w_{3ij} * w_{2kl} * w_{1mn} * \text{features}$
  - If w_{3ij} or w_{2kl} is small
    the gradient "signal" telling w_{1mn} in which way to change
    does not reach w_{1mn}

- Solution: link w_{1mn} to *z* directly – via a *skip connection*



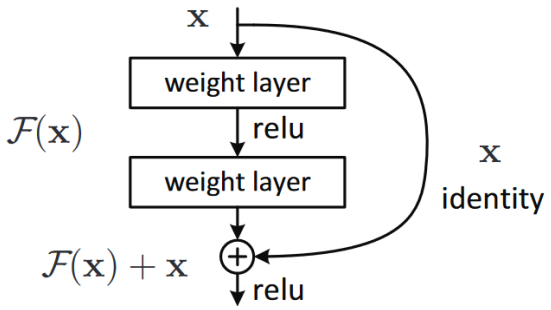
Skip connections

- Solution: link w_{1mn} to z directly – via a skip connection

- Without skip connection:
  - We see terms like $z=w_{3ij}*w_{2kl}*w_{1mn}*input$
  - The derivative of z over w_{1mn} is $w_{3ij}* w_{2kl} *input$

- With skip connection:
  - Connect $w_{1mn}*x$ directly to z, in addition to going through w_3 and w_2
  - We have $z=w_{3ij}*w_{2kl}*w_{1mn}*input + w_{1mn}*input$

    $z=(w_{3ij}*w_{2kl}+1)*w_{1mn}*input$
  - The derivative of z over w_{1mn} is $w_{3ij}* w_{2kl} *input$ + input

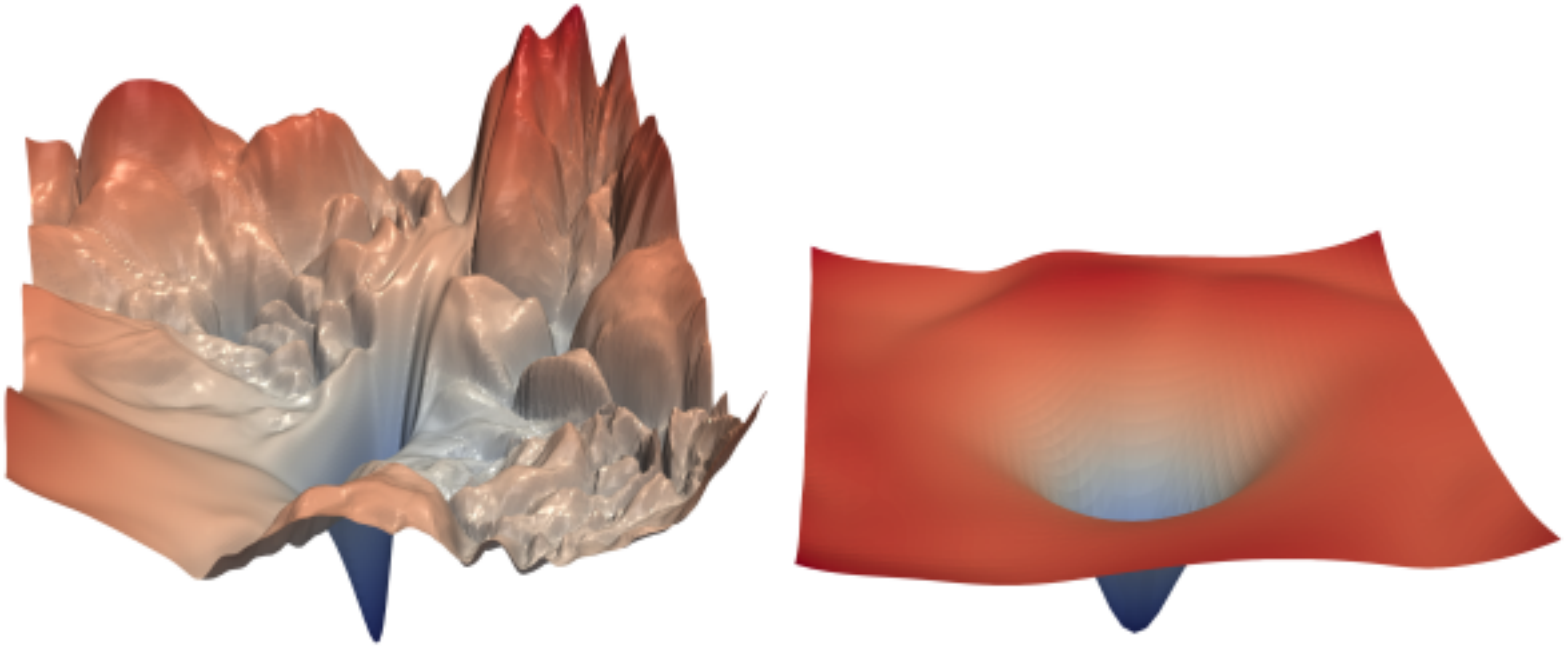


ResNet

- Adding skip connections makes the loss landscape smoother

$$z=(w_{3ij}*w_{2kl}+1)*w_{1mn}*x$$

less reliance just on multiplication

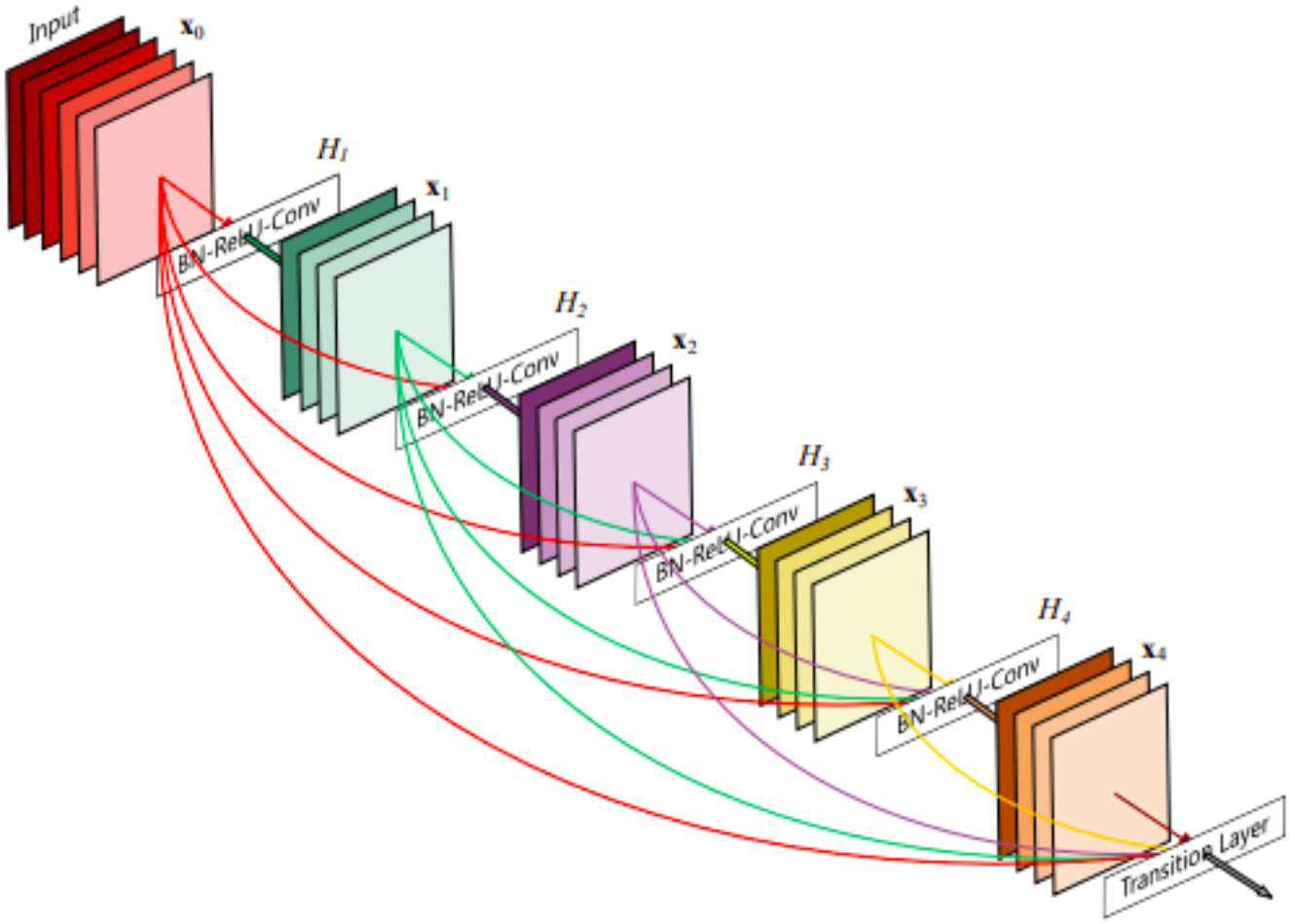


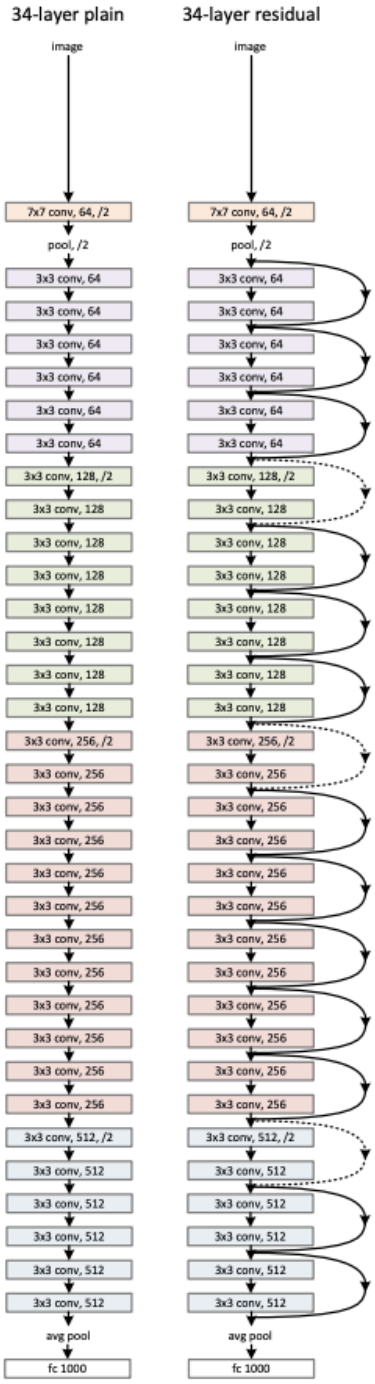
(a) without skip connections

(b) with skip connections

Skip connections

- Modern network architectures often involve many skip connections

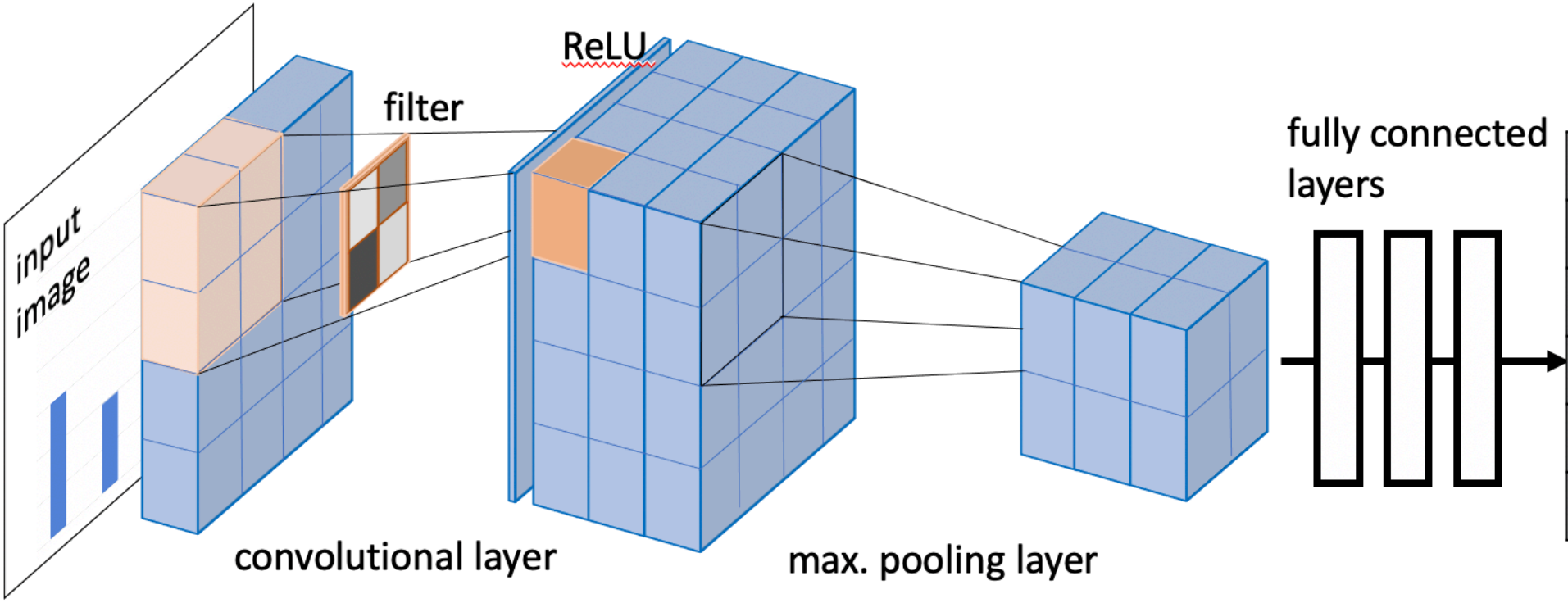




U-Net architecture

ConvNet:

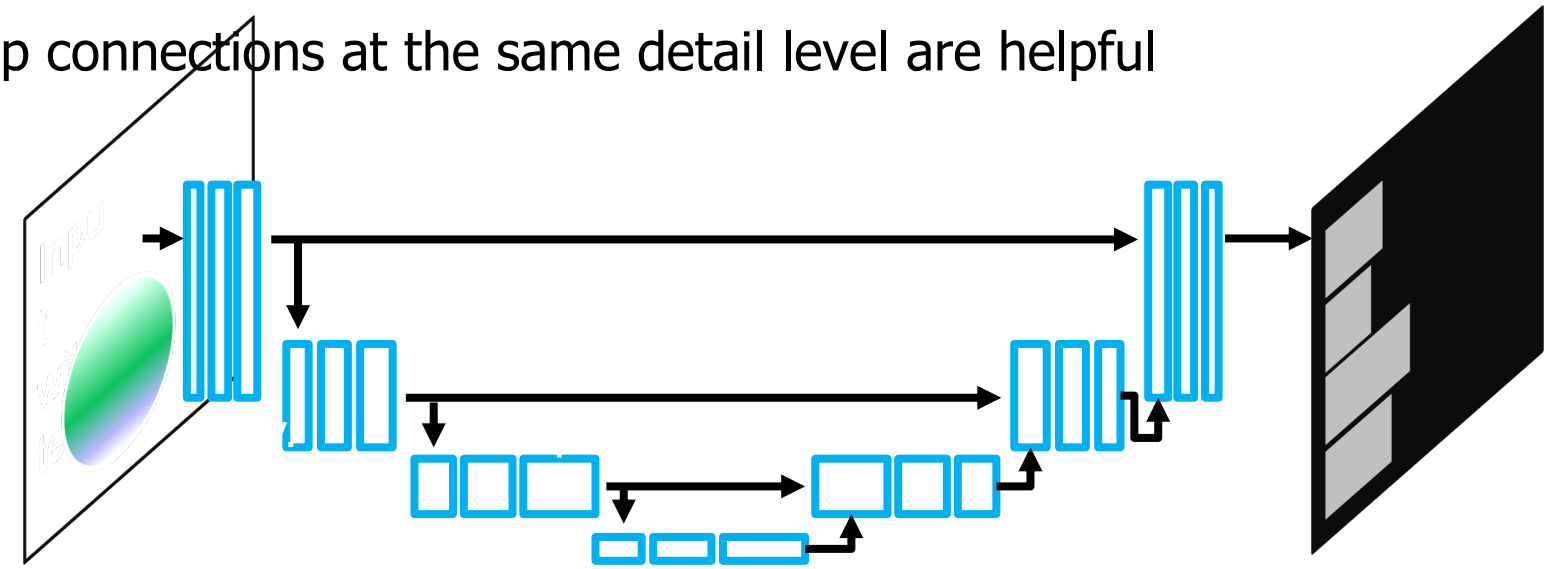
downsampling via maxpooling



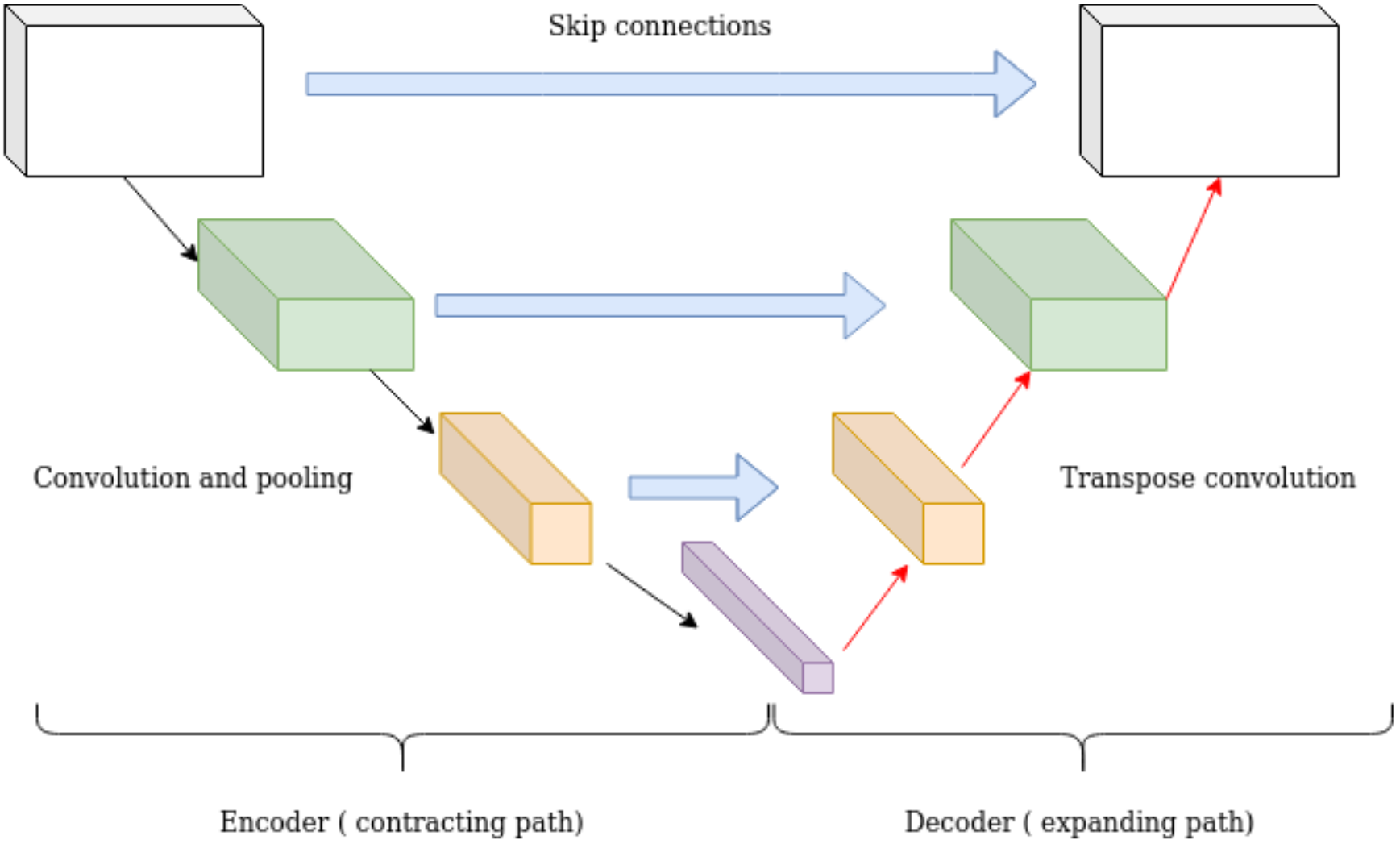
U-Net architecture

- Skip connections are used in image-to-image networks

- E.g. U-Net architecture
  - Convnet downsampling (via max pooling)
    - From detail to coarse-grained
  - Upsampling
    - Generate detail from coarse-grained
  - Skip connections at the same detail level are helpful



U-Net architecture



Skip connections - summary

- Skip connection can help with optimization
- They can also help with preserving details (information) from earlier layers, e.g. in U-Net architecture
- Networks with skip connections are often called residual networks (ResNets)
  - They only need to learn the residual (the "delta") f(x)-x not the whole f(x) from scratch