3050571 Practical Clin Data Sci

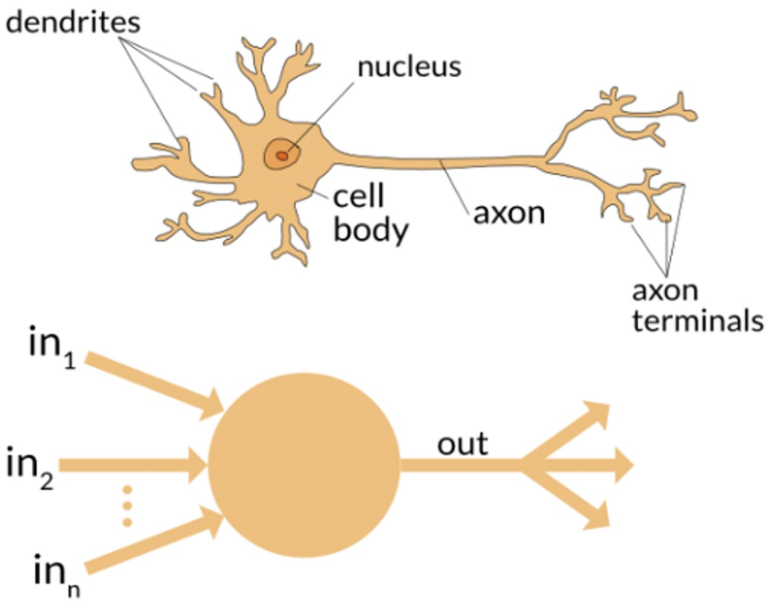
Session 14: Artificial neural network designs

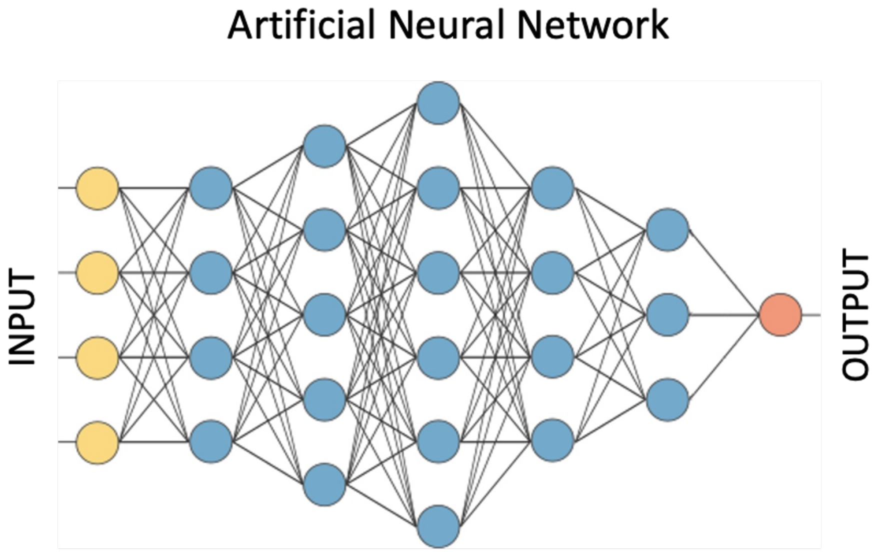
February 29, 2024

Sira Sriswasdi, PhD

- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

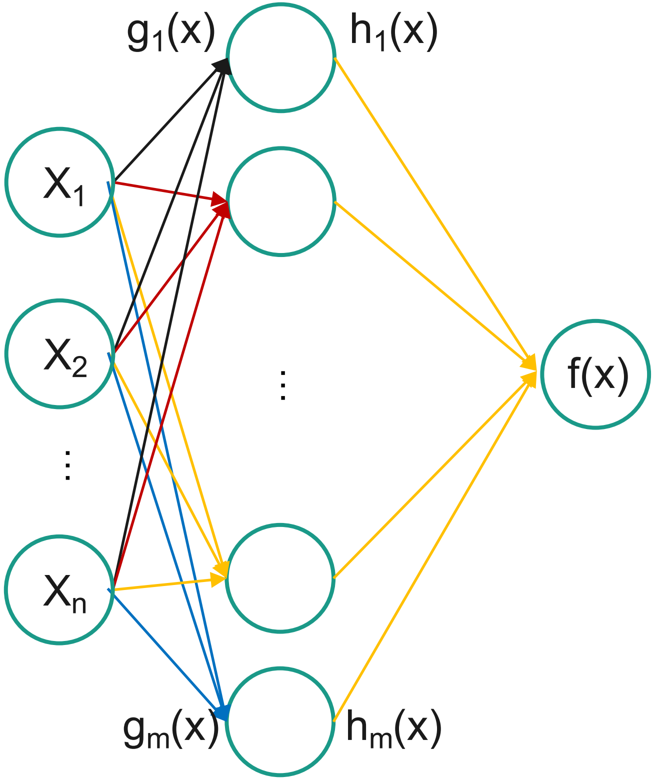
Artificial neural network





- Network of simple computation nodes: $out = f(w_1 in_1 + w_2 in_2 + \ldots + w_n in_n)$

Calculations inside neural nentwork



Linear neuron input

- $g_1(x) = w_{1,1}x_1 + \cdots + w_{1,n}x_n$
- $g_m(x) = w_{m,1}x_1 + \cdots + w_{m,n}x_n$

Sigmoid activation

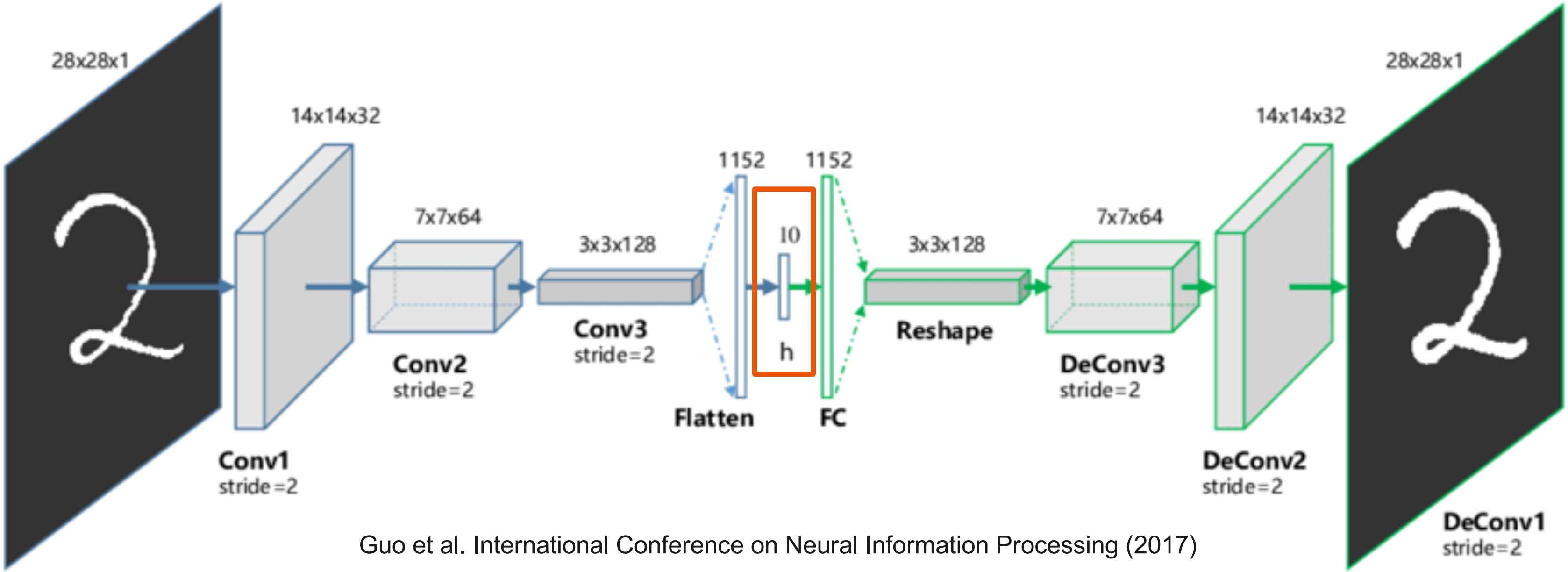
- $h_1(x) = \frac{1}{1+e^{-g_1(x)}}$
- $h_m(x) = \frac{1}{1+e^{-g_m(x)}}$

Linear aggregated output

- $f(x) = u_1h_1(x) + \cdots + u_mh_m(x)$

Autoencoder

Representation learning via self-reconstruction



Guo et al. International Conference on Neural Information Processing (2017)

- Similar to dimensionality reduction

Denoising autoencoder

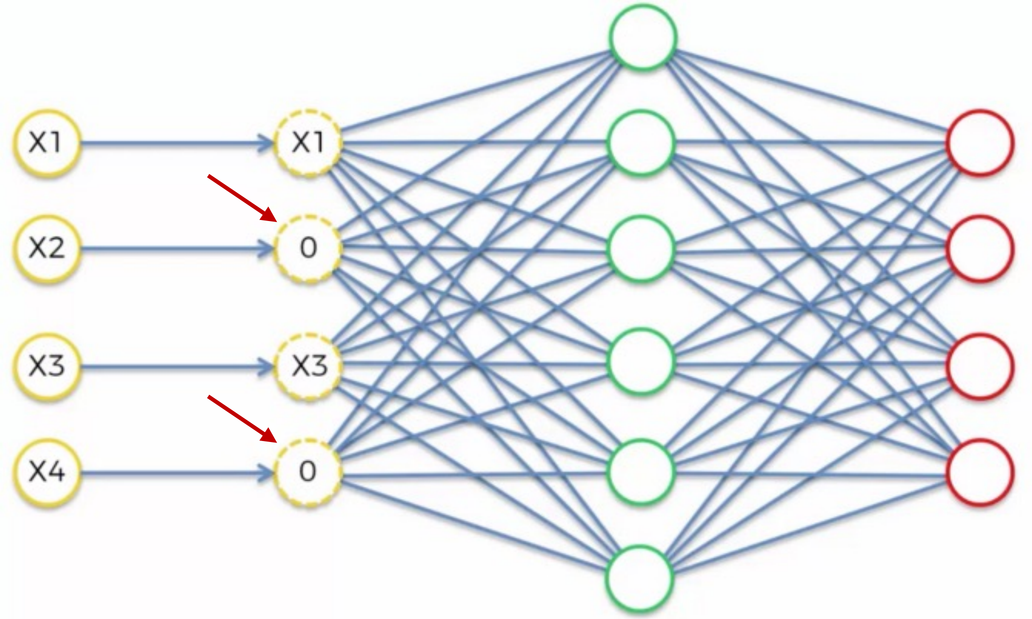


Image from towardsdatascience.com/denoising-autoencoders-explained-dbb82467fc2

- Randomly set some inputs to zero → robust representation

Variational autoencoder (VAE)

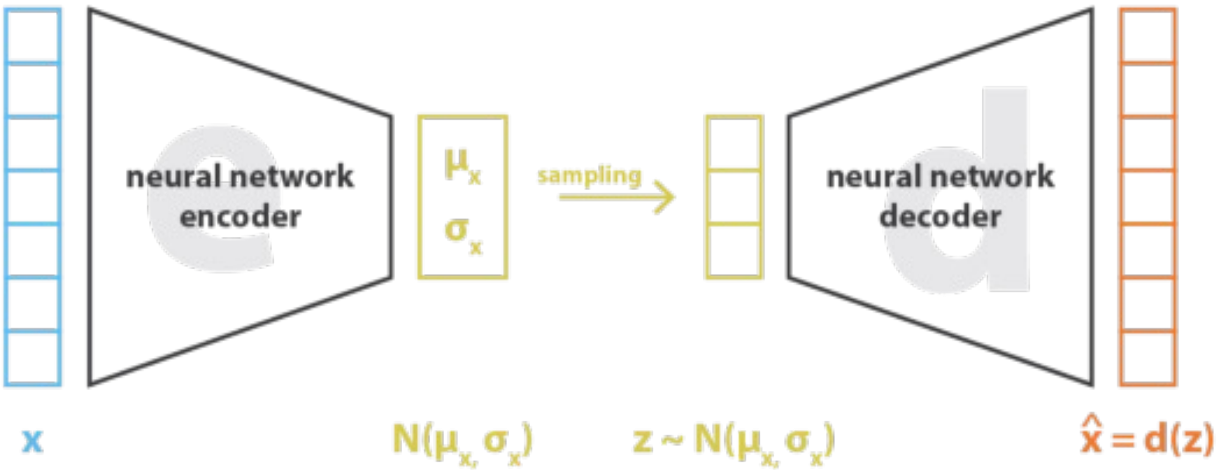
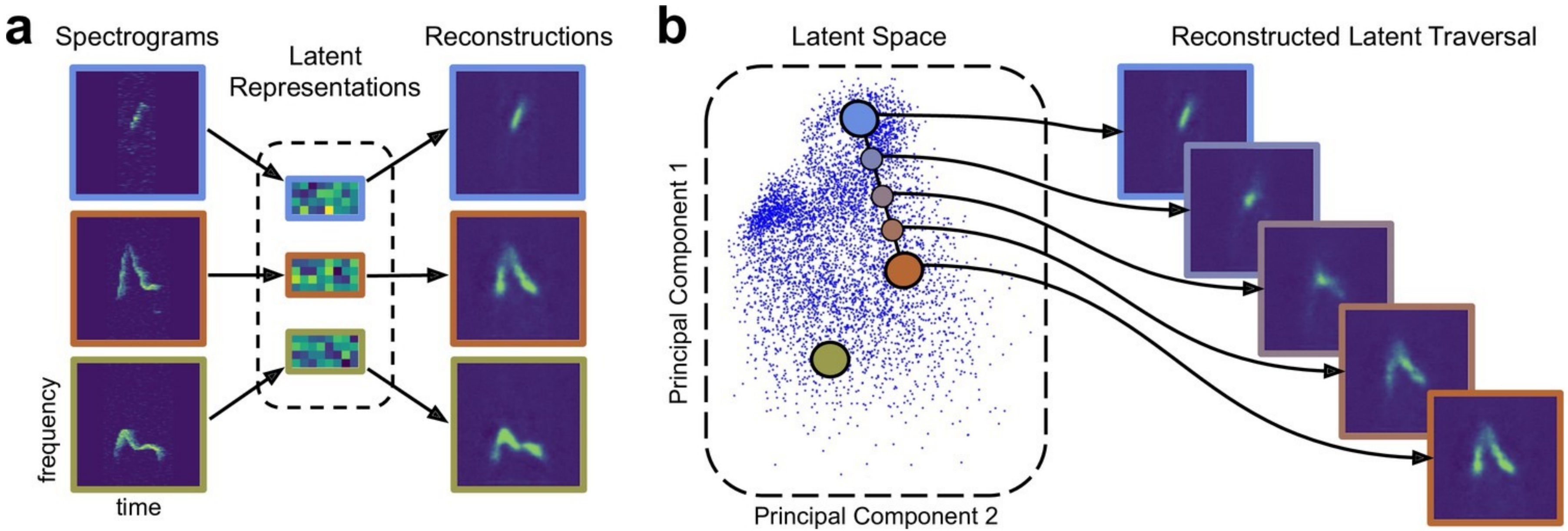


Image from www.jeremyjordan.me/variational-autoencoders/

- Learned representation = parameters for distribution
- Decoder is robust to small changes in the representation
  - Smooth representation space

VAE generates smoother representation space

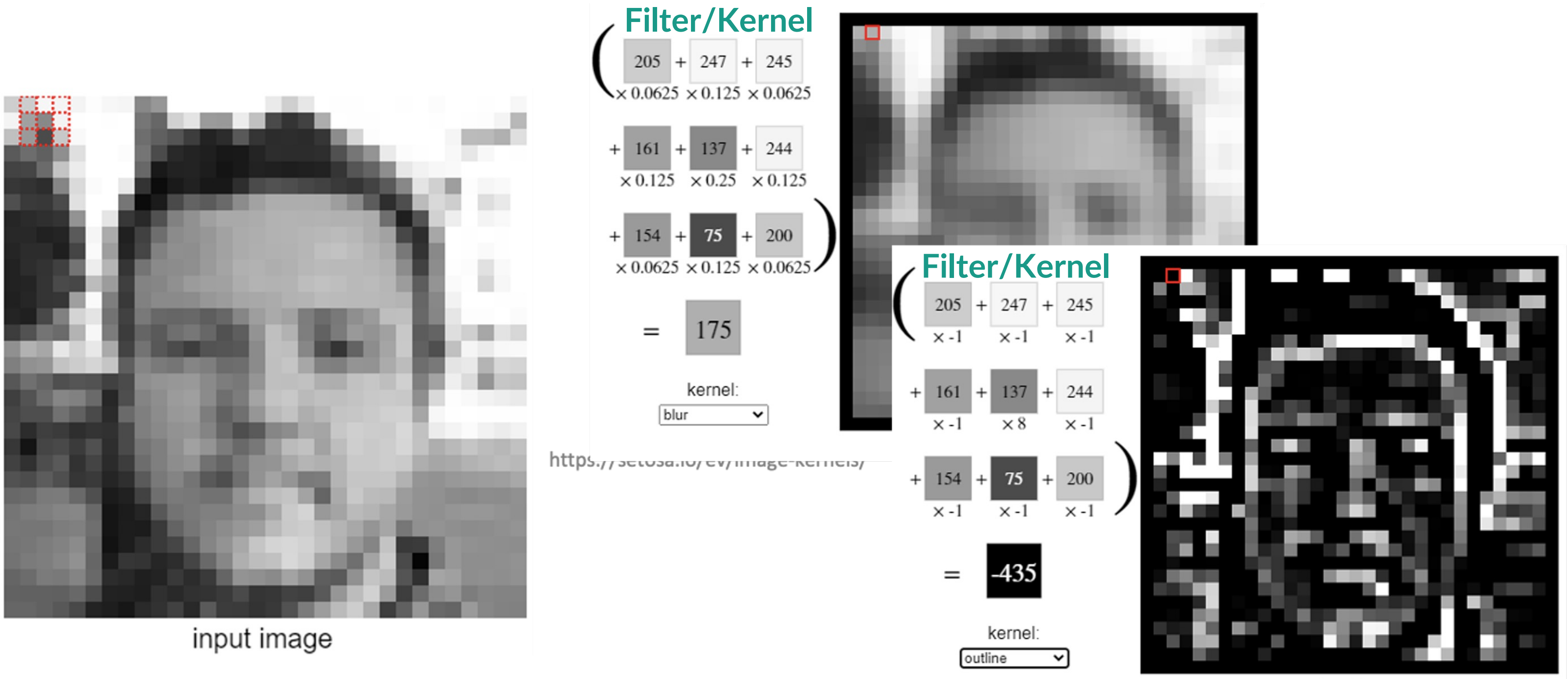


Doffinet, J. et al. eLife 67855 (2021)

- VAE learn representation distribution, not just individual vectors

Convolutional neural network

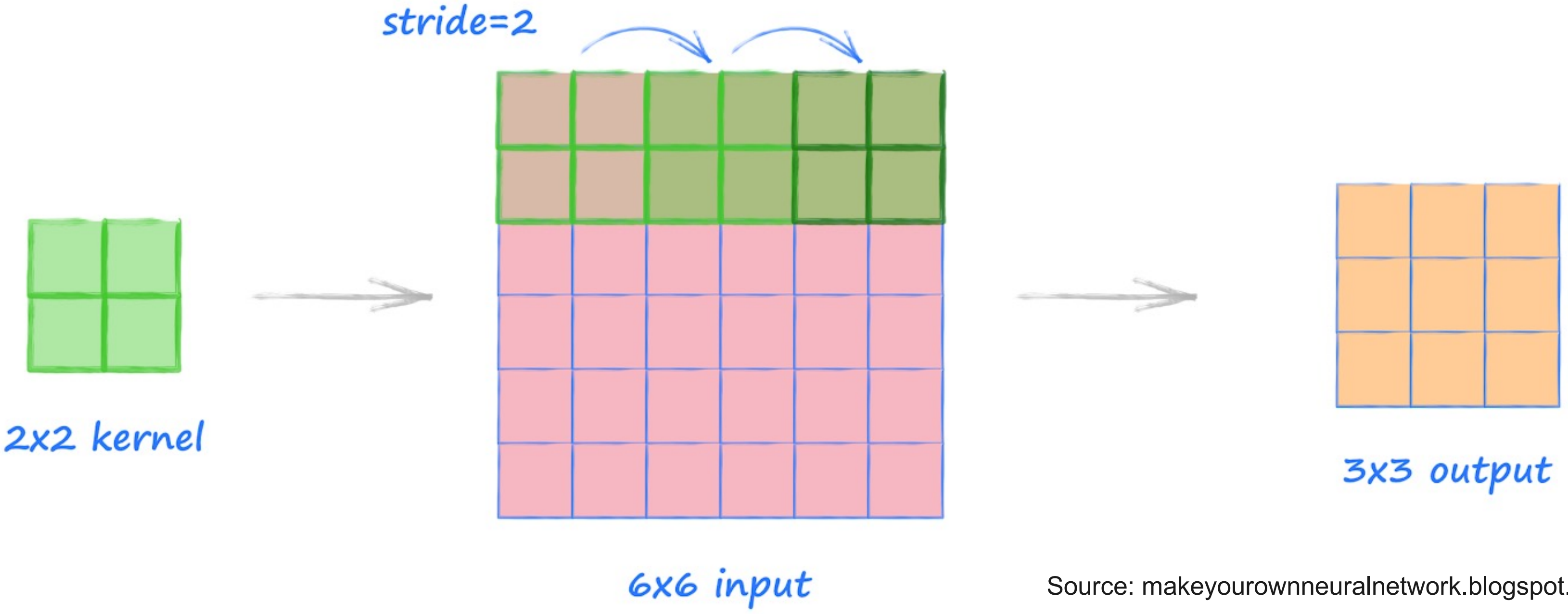
Extracting contextual pattern with filter



input image

output image

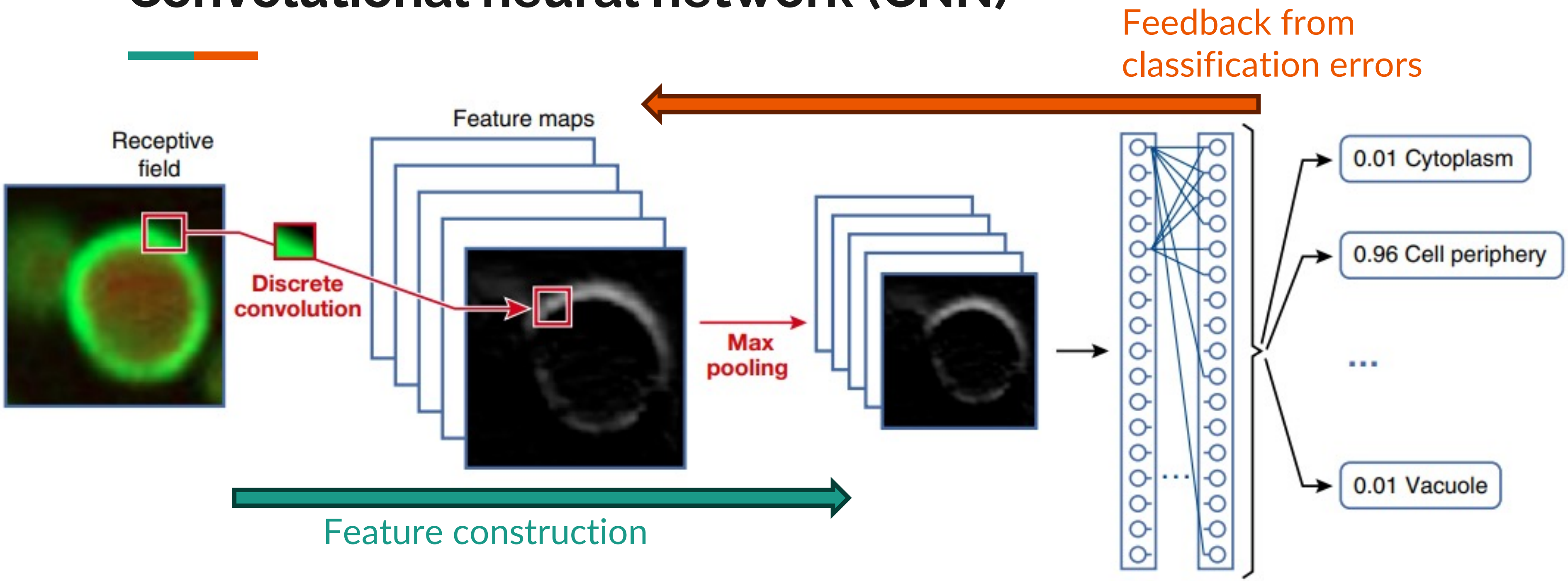
Convolutional operation



Source: makeyourownneuralnetwork.blogspot.com

- Linear combination of values in nearby pixels – applied throughout

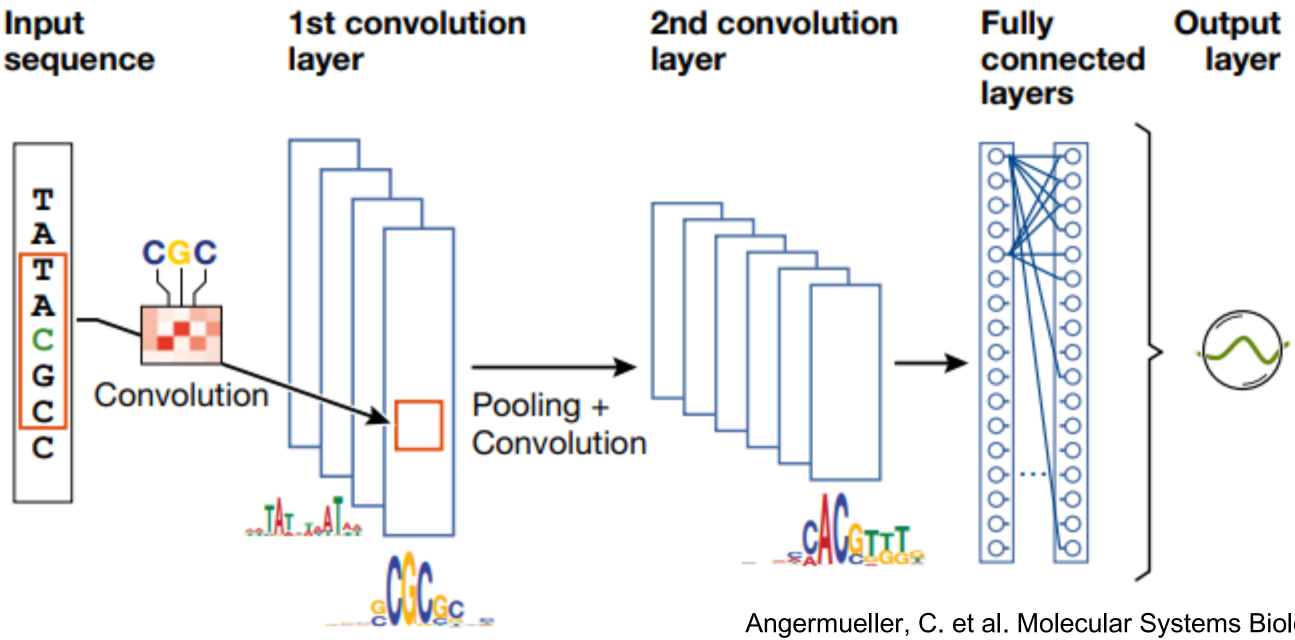
Convolutional neural network (CNN)



- Instead of using human-define filters to extract contextual pattern, CNN learns the best filters from the data

Angermueller, C. et al. Molecular Systems Biology 12:878 (2016)

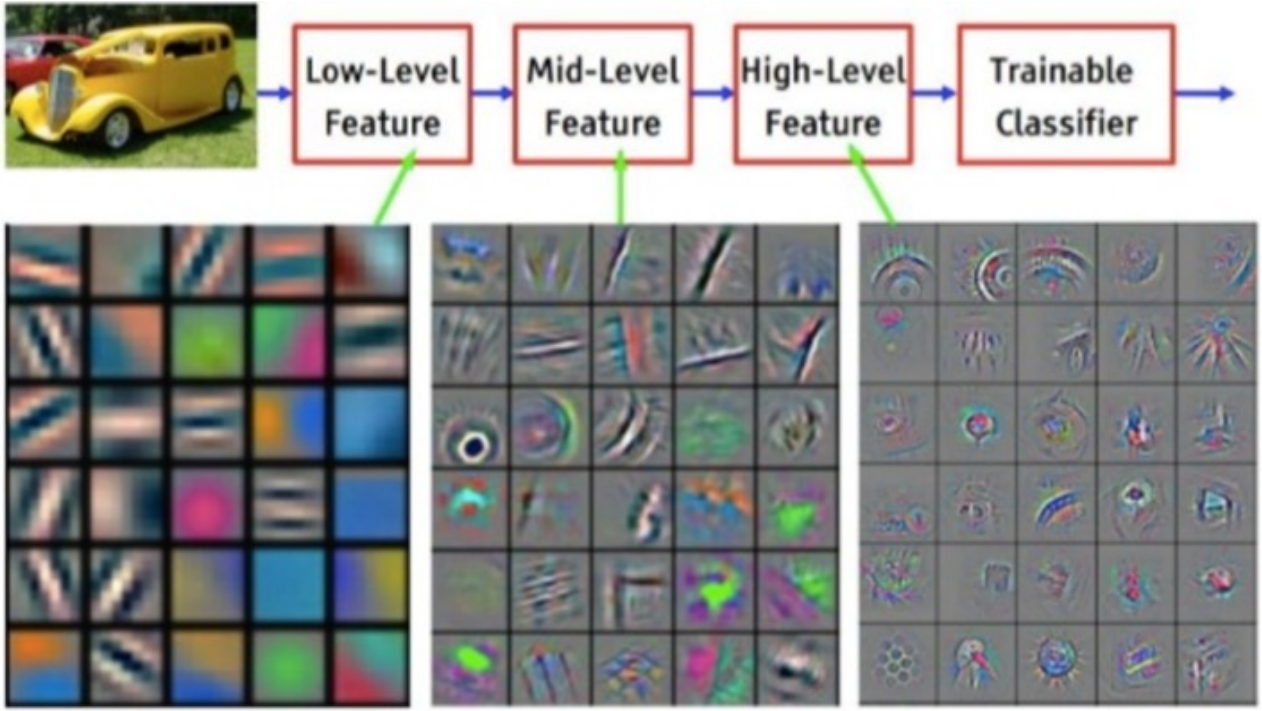
Convolution for DNA sequences



Angermueller, C. et al. Molecular Systems Biology 12:878 (2016)

- Motif = contextual pattern on DNA sequence

Hierarchical feature assembly inside CNN



Source: Zeiler and Fergus (2013)

Some CNN designs

Vanishing and exploding gradient problems

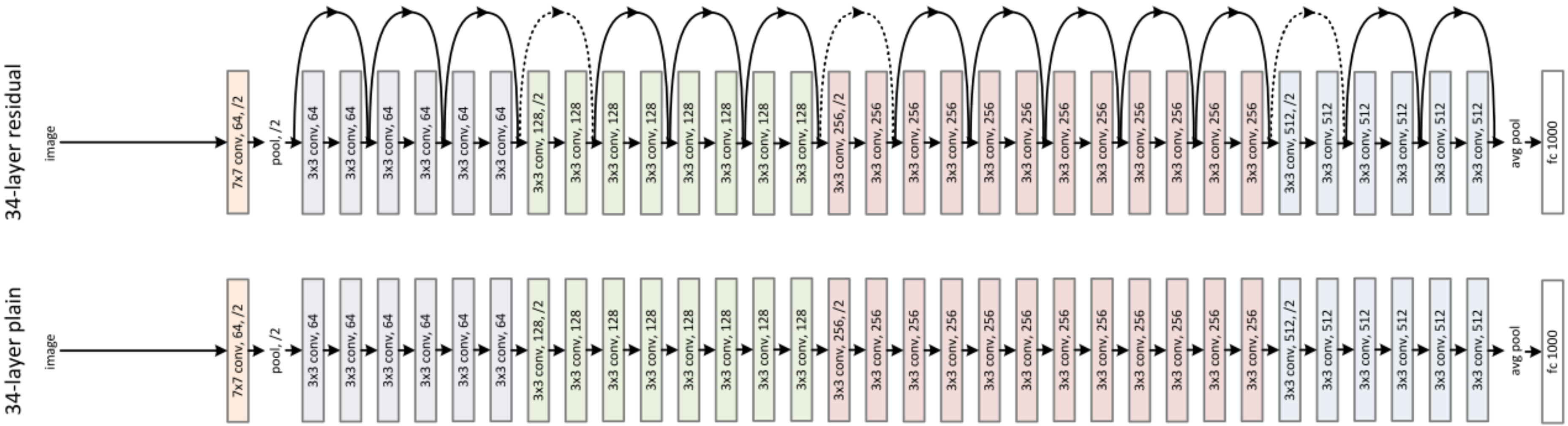
Gradient: $\frac{\delta L}{\delta w_{i,j}} = \frac{\delta L}{\delta f}\frac{\delta f}{\delta h_i}\frac{\delta h_i}{\delta g_i}\frac{\delta g_i}{\delta w_{i,j}} = (f(x) - y) \cdot u_i \cdot g_i(\mathrm{x})(1 - g_i(\mathrm{x}))\ x_j$

The number of multiplicative terms scales with the number of layers

What would happen if all values are << 1 or >> 1?

- Gradient became very small → No weight update
- Gradient became very large → Unstable

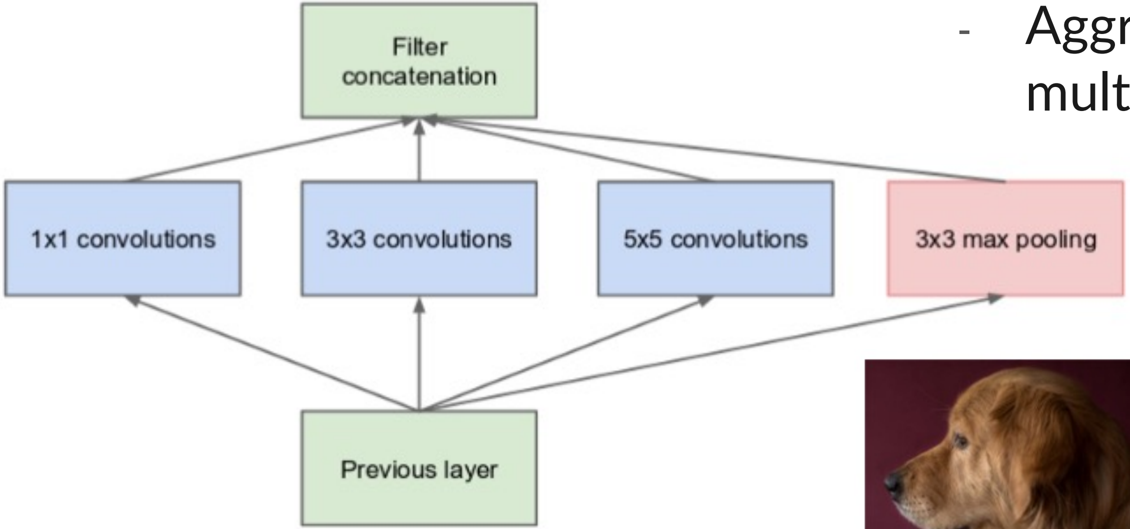
Residual network (ResNet)



Source: medium.com

- Adding skip connections jumping over blocks of convolutional layers
- Reduce the number of terms in gradient of early weights

Inception = multi-resolution layer

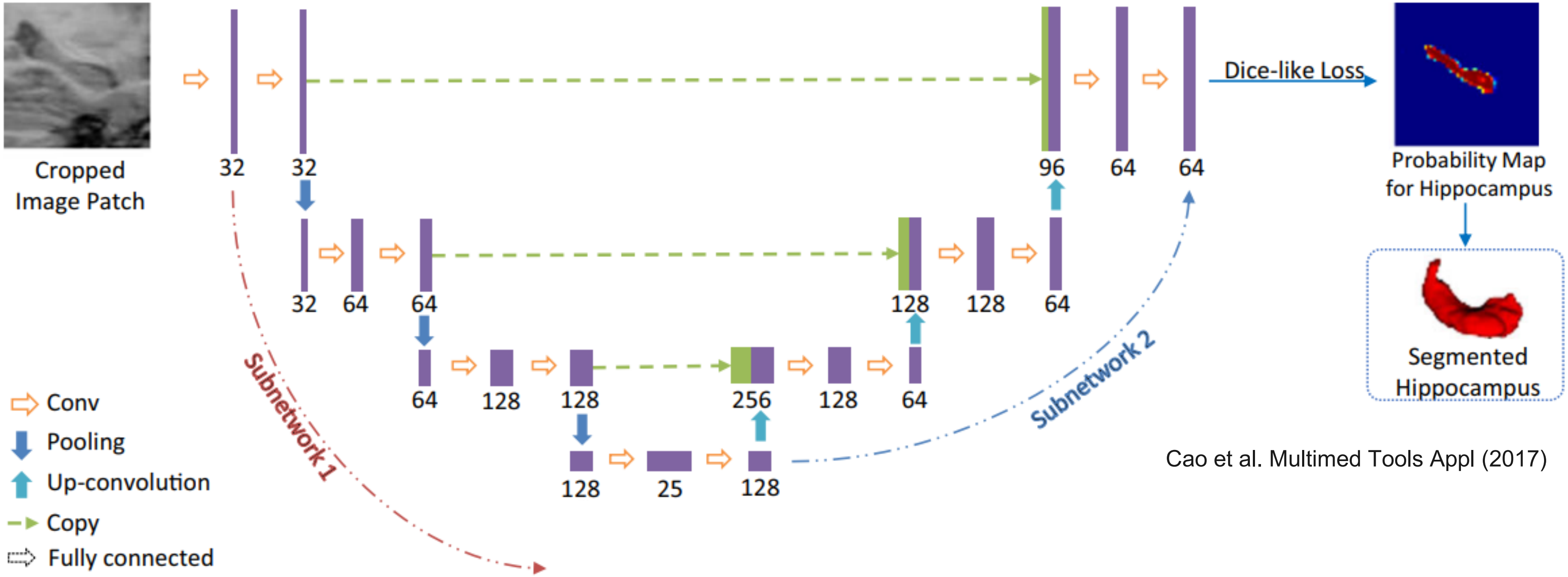


- Aggregate information from multiple resolutions together

From left: A dog occupying most of the image, a dog occupying a part of it, and a dog occupying very little space (Images obtained from Unsplash).

- Convolution with multiple filter sizes per layer

U Net = producing image from image

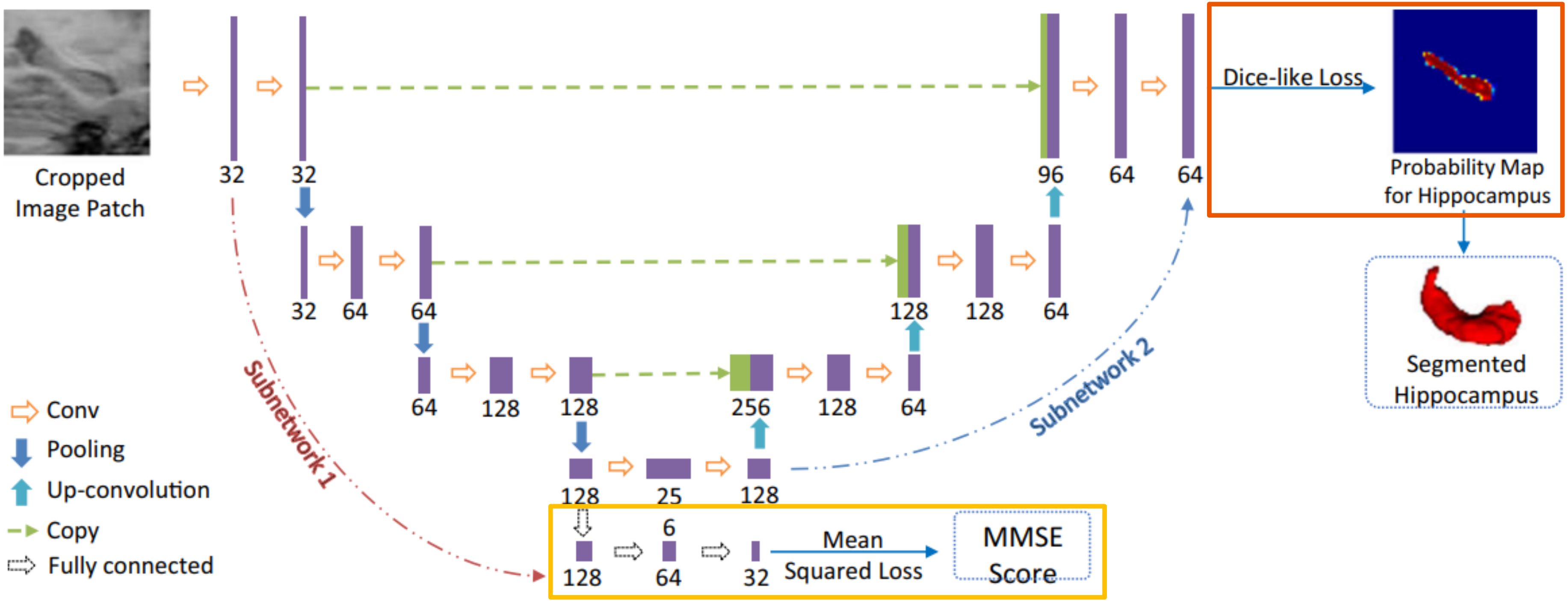


Cao et al. Multimed Tools Appl (2017)

- Make prediction for every pixel → output size = input size

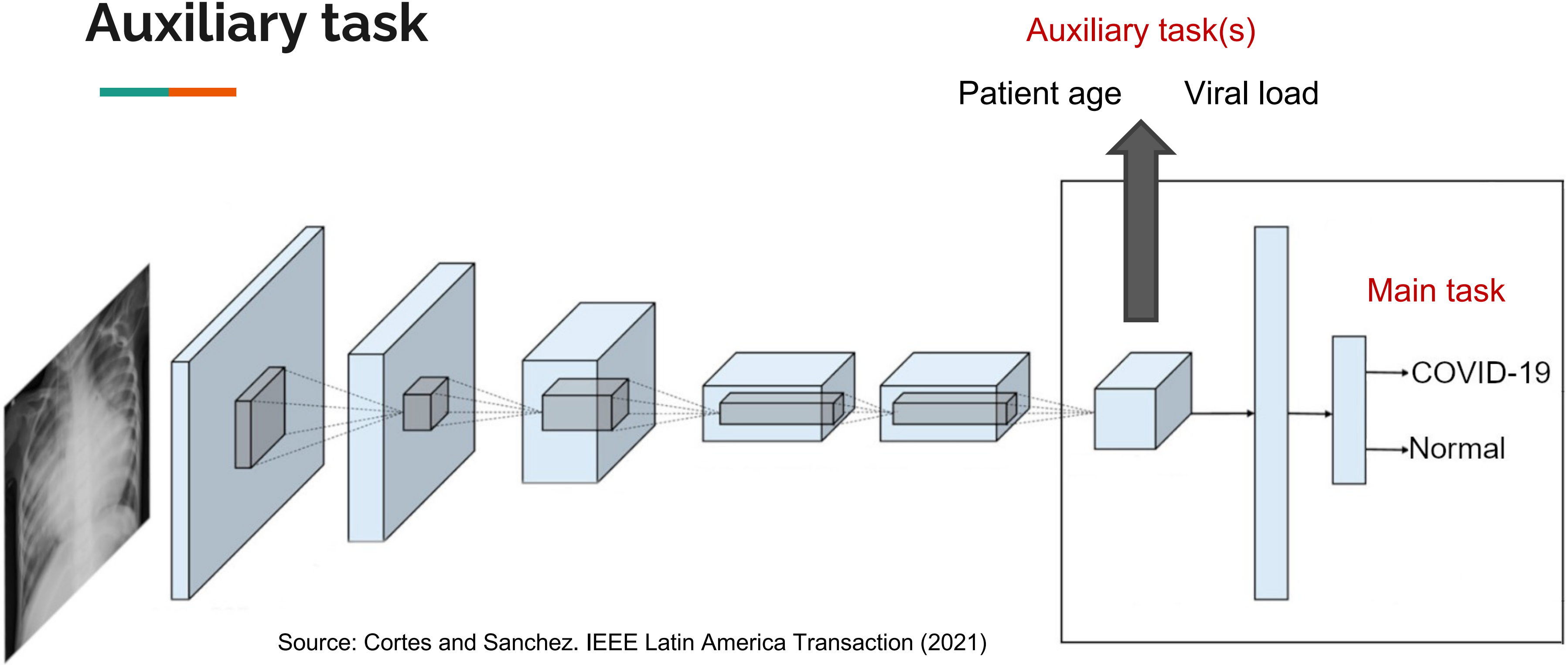
Multitasking

Simultaneous segmentation & classification



- Combine gradients from both tasks

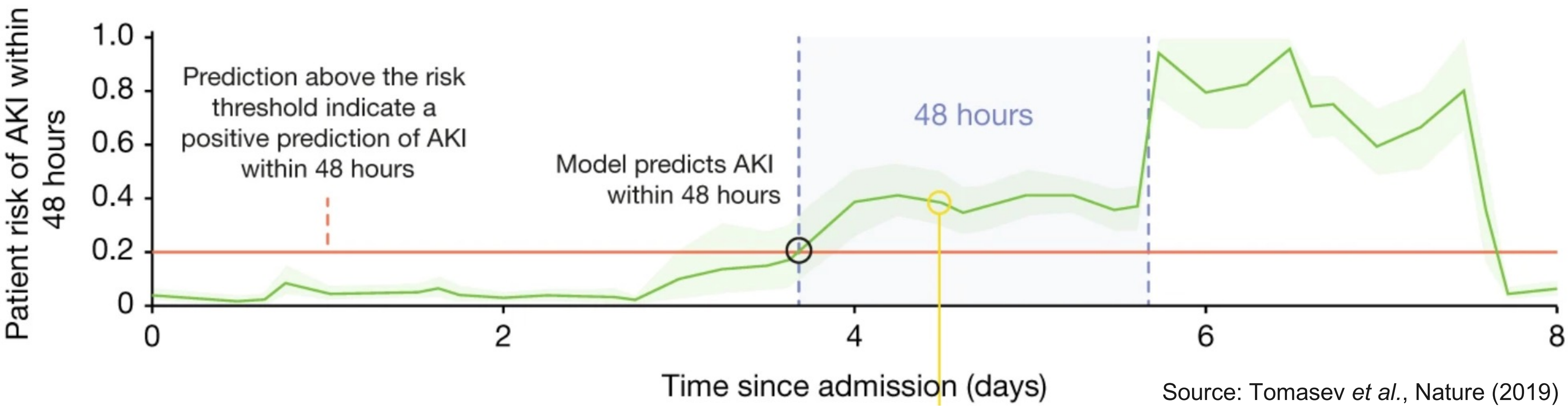
Auxiliary task



Source: Cortes and Sanchez. IEEE Latin America Transaction (2021)

- Encourage the learned representation to include more information

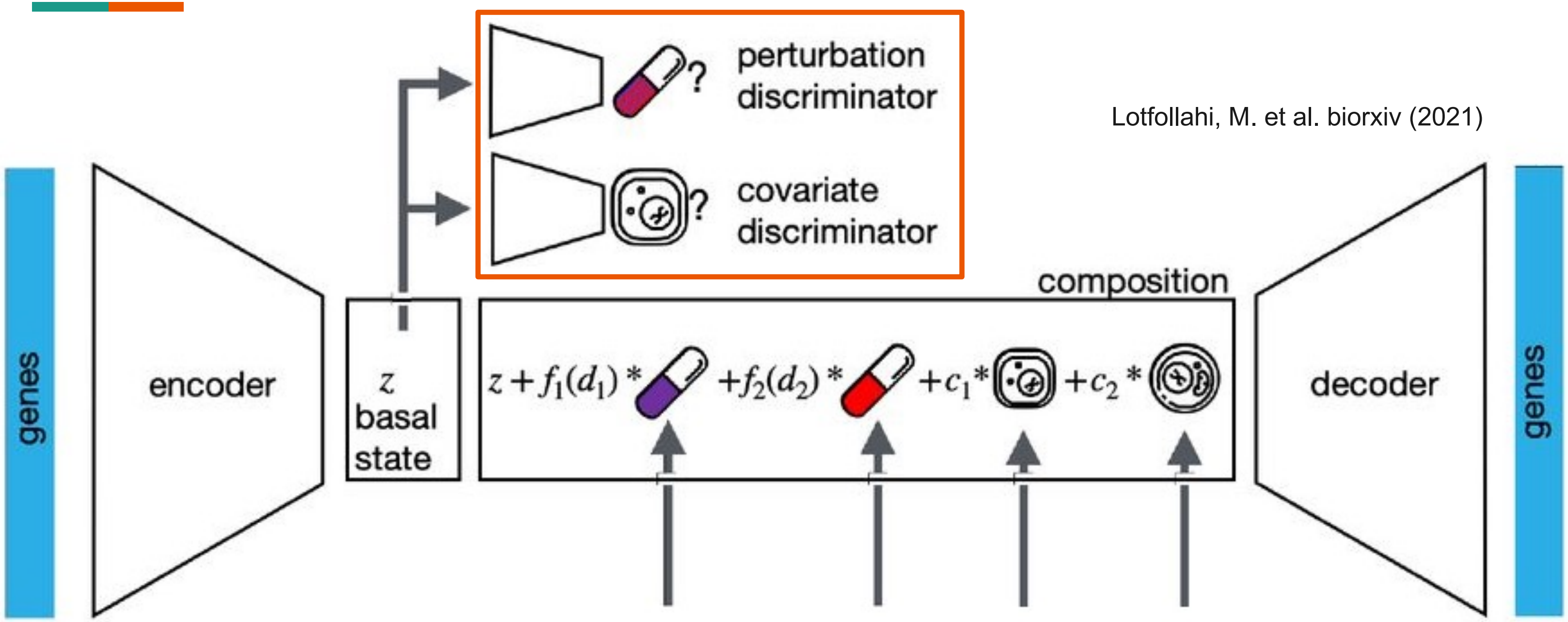
Acute kidney injury prediction



Source: Tomasev *et al.*, Nature (2019)

- **Main task**: Occurrence of acute kidney injury within 48 hours
- **Auxiliary tasks**: Maximal values of 7 key lab tests within 48 hours
    - Provide more feedback on what the model gets wrong

Decoupling / debiasing



Lotfollahi, M. et al. biorxiv (2021)

- Deconvolute cell basal state from perturbation and covariate
- Update weights in the opposite direction of gradient

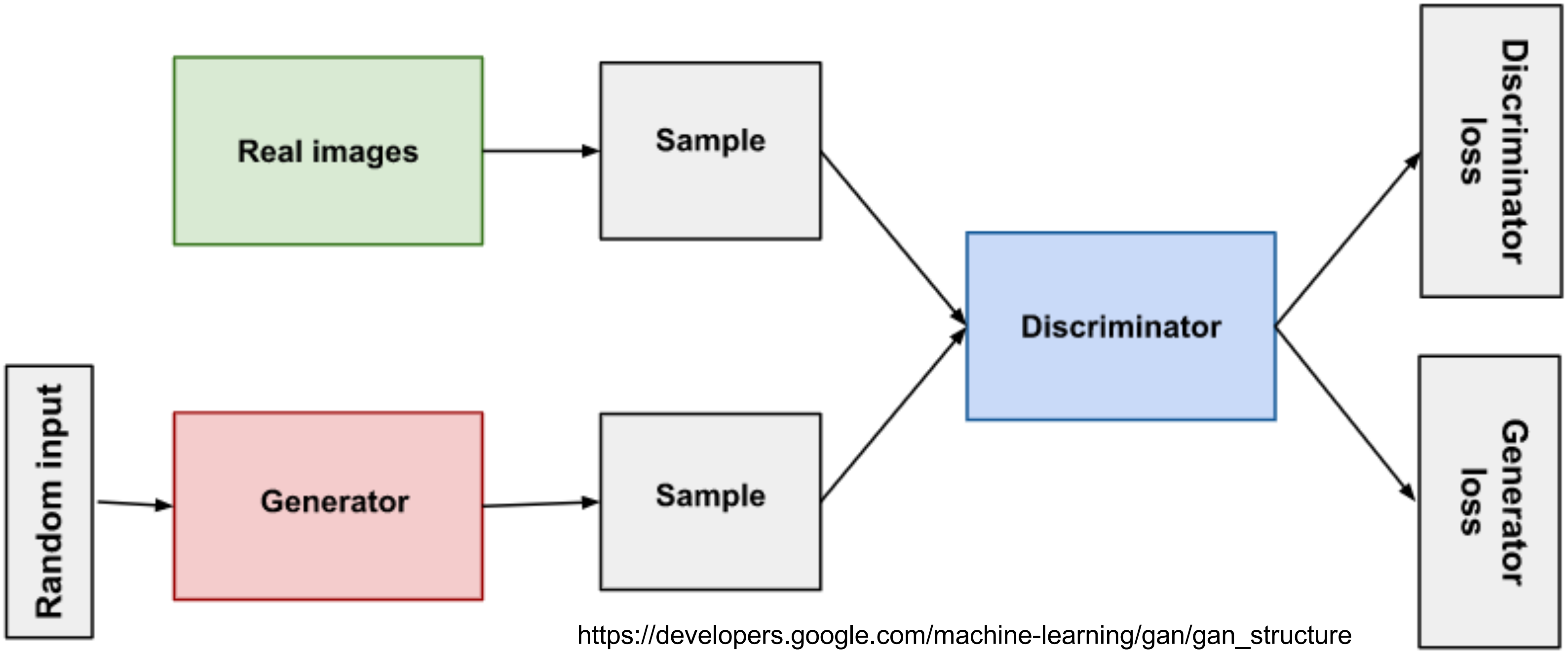
Generative model

Why generative model?

https://developers.google.com/machine-learning/gan/gan_structure

- Models that generate realistic data can tell us about the underlying mechanisms of the system

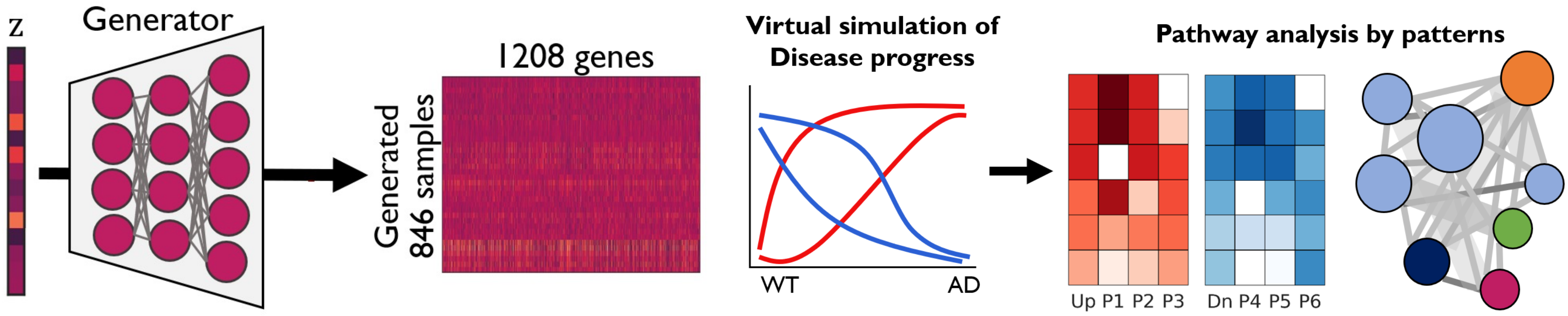
Generative adversarial network (GAN)



https://developers.google.com/machine-learning/gan/gan_structure

- Simultaneous training of generator and discriminator

Knowledge from simulated data

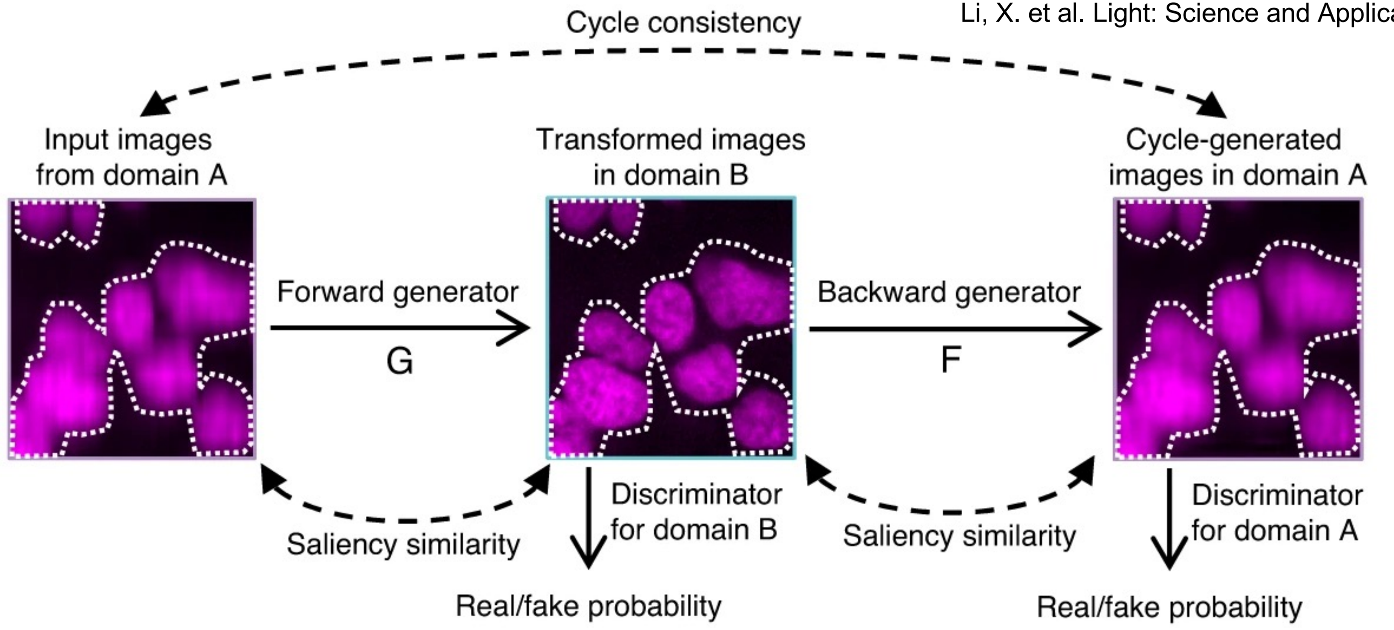


Park, J. et al. PLoS Computational Biology 16:e1008099 (2020)

- Train a generator with data from small-scale experiment
- Simulate time-course gene expression profiles
- Perform usual bioinformatics analyses to infer biological knowledge

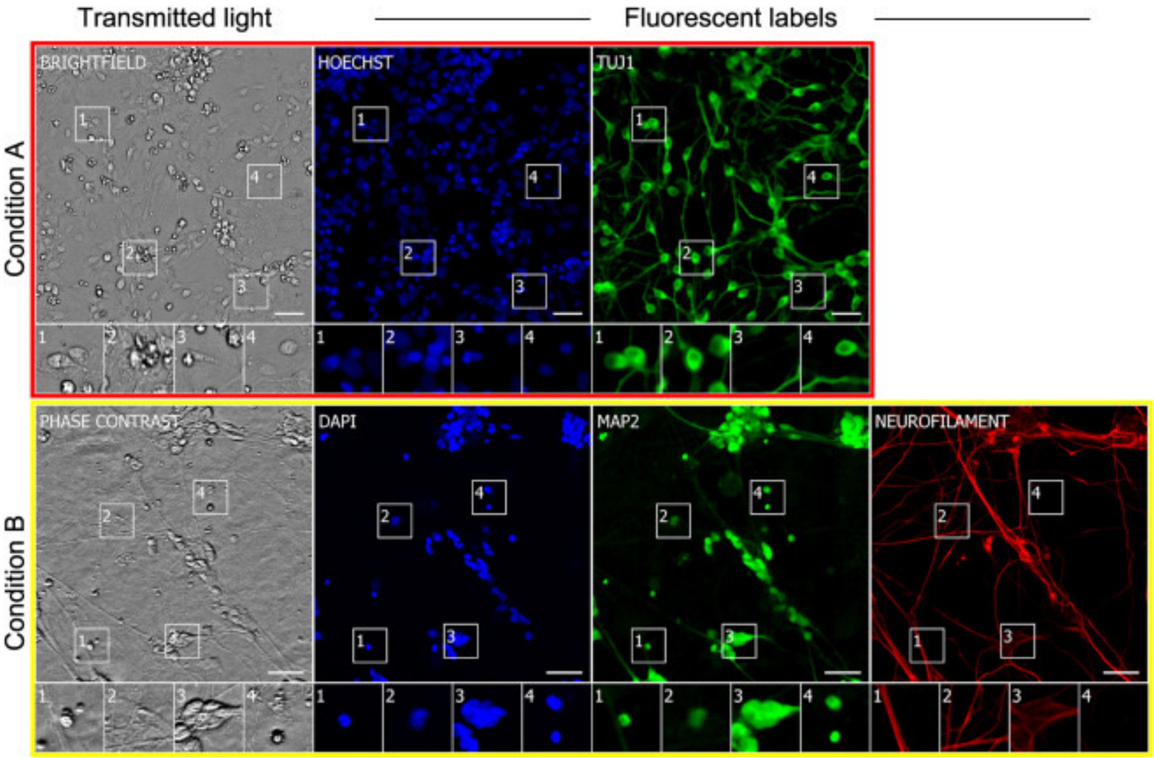
Cycle GAN for transforming image

Li, X. et al. Light: Science and Applications10:44 (2021)

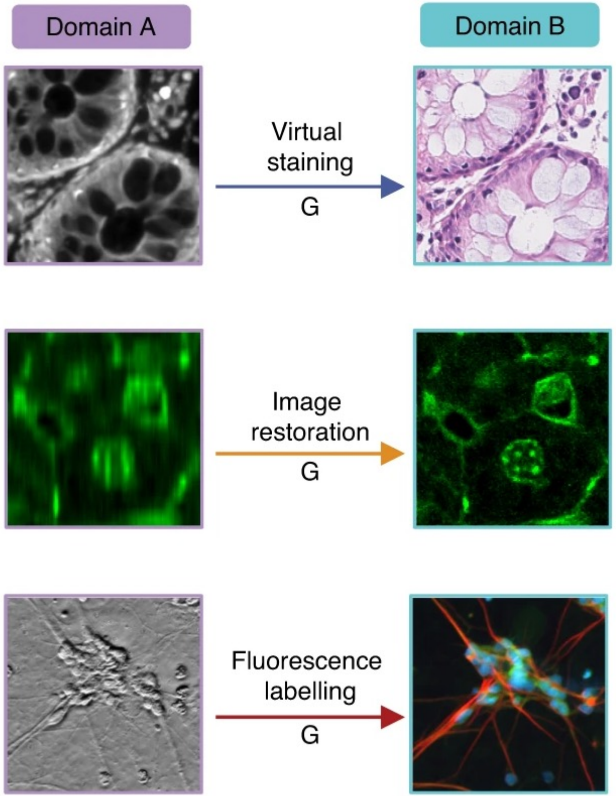


- Generate sharpened image from blurry image and back

Virtual staining



Christiansen, E.M. et al. Cell 173:792-803.e19 (2018)

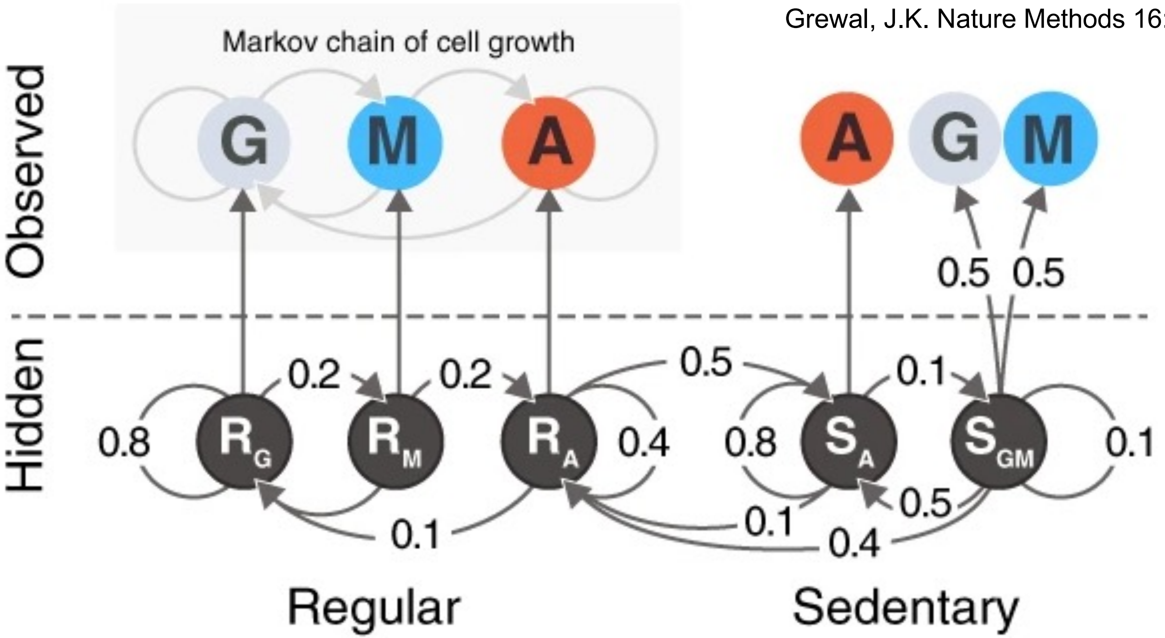


Li, X. et al. Light: Science and Applications10:44 (2021)

Recurrent neural network

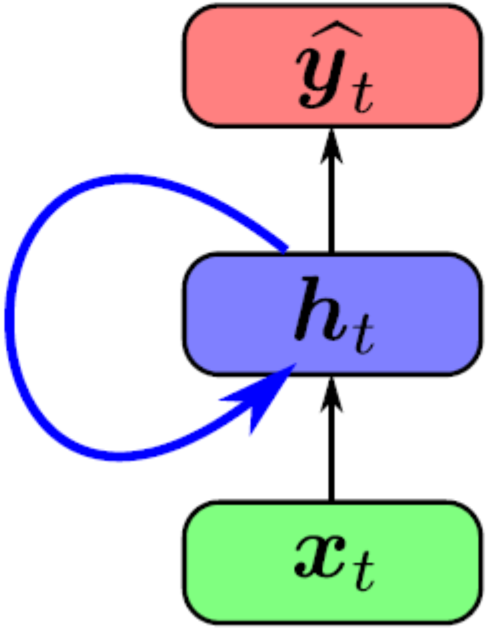
Hidden Markov Model

Grewal, J.K. Nature Methods 16:795-796 (2019)



- Sequence of observations, each generated from a model

Recurrent neural network



Output of a hidden layer

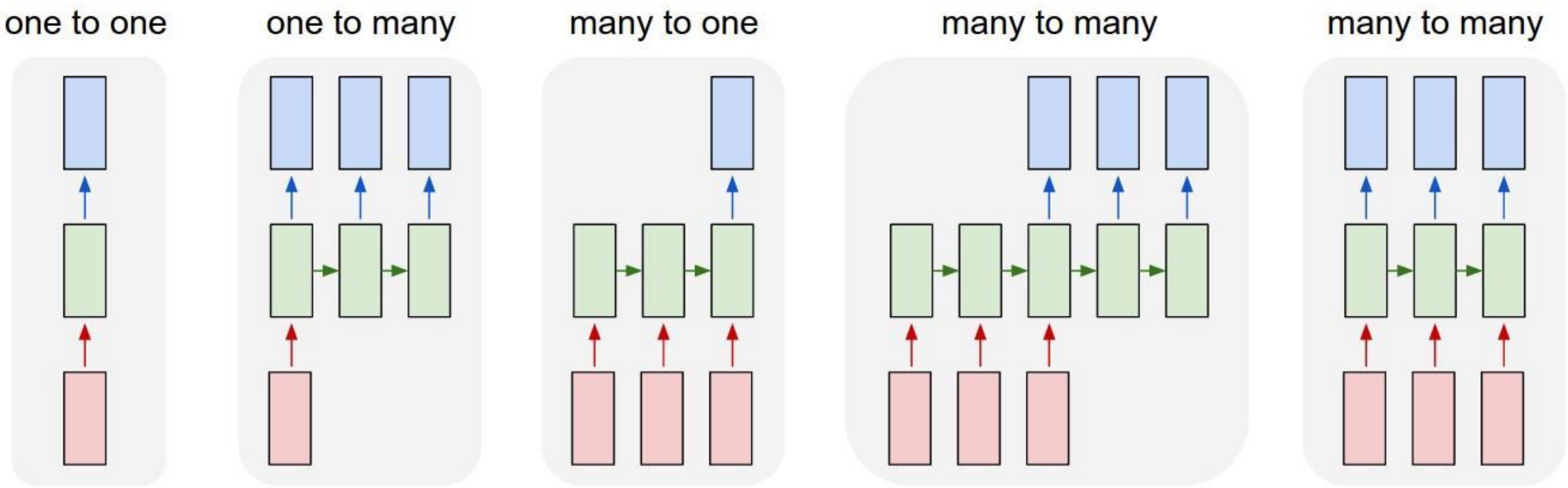
Variable-length input

Shared weights!

$$h_1 = f(\boldsymbol{u} \cdot x_1 + \boldsymbol{v} \cdot h_0 + c)$$
$$h_2 = f(\boldsymbol{u} \cdot x_2 + \boldsymbol{v} \cdot h_1 + c)$$
$$\cdots$$
$$h_t = f(\boldsymbol{u} \cdot x_t + \boldsymbol{v} \cdot h_{t-1} + c)$$
$$\hat{y}_t = \boldsymbol{w} \cdot h_t + b$$

- Reuse a single layer (weights) over time with different input

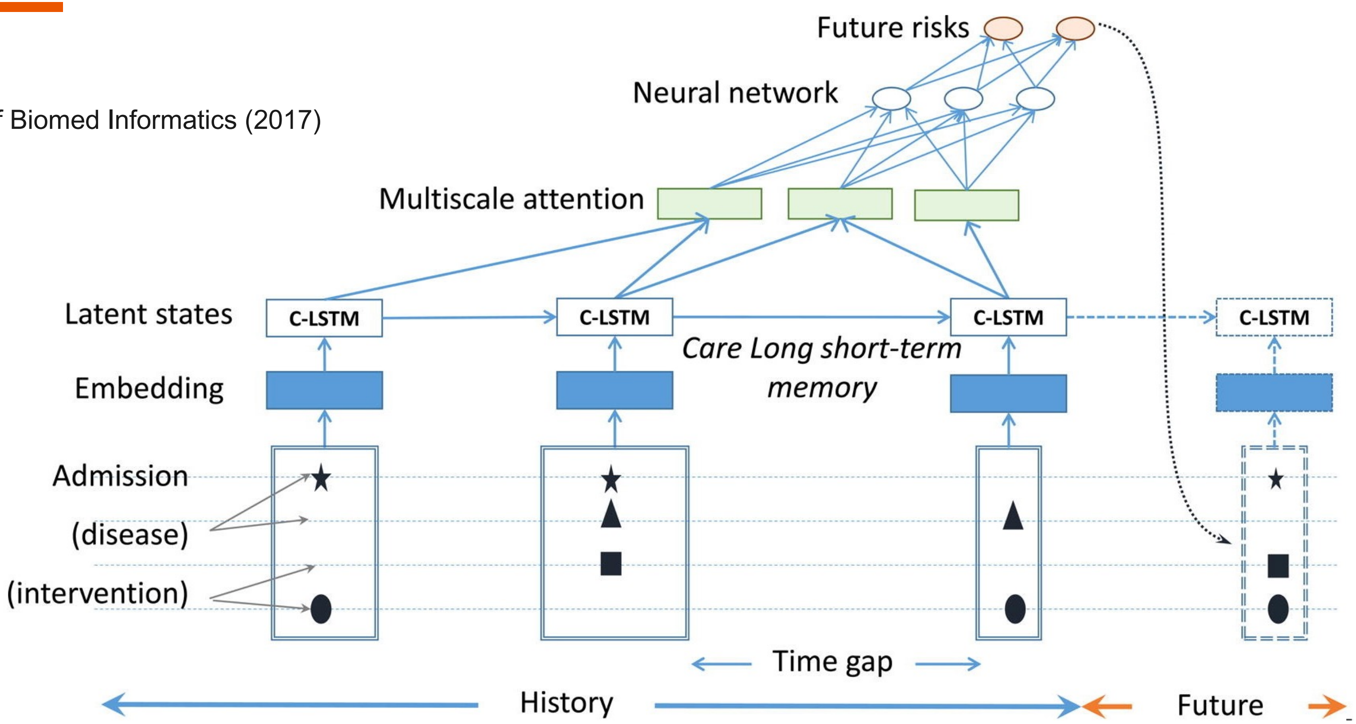
Sequence-to-sequence capability



- Handle variable input and output

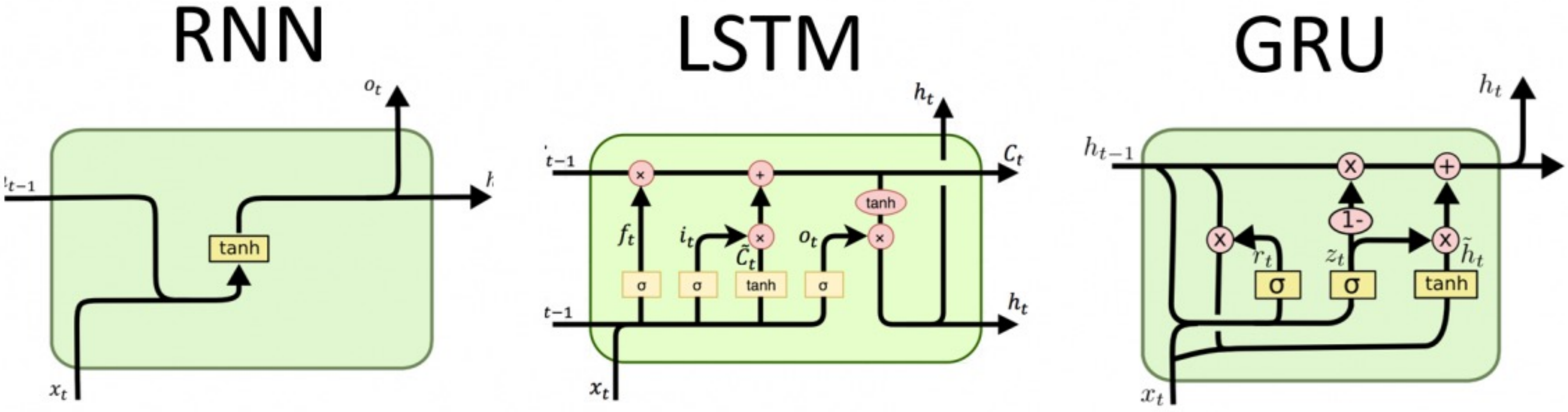
RNN on medical history

Pham et al. J of Biomed Informatics (2017)



- Aggregate information across time to make prediction

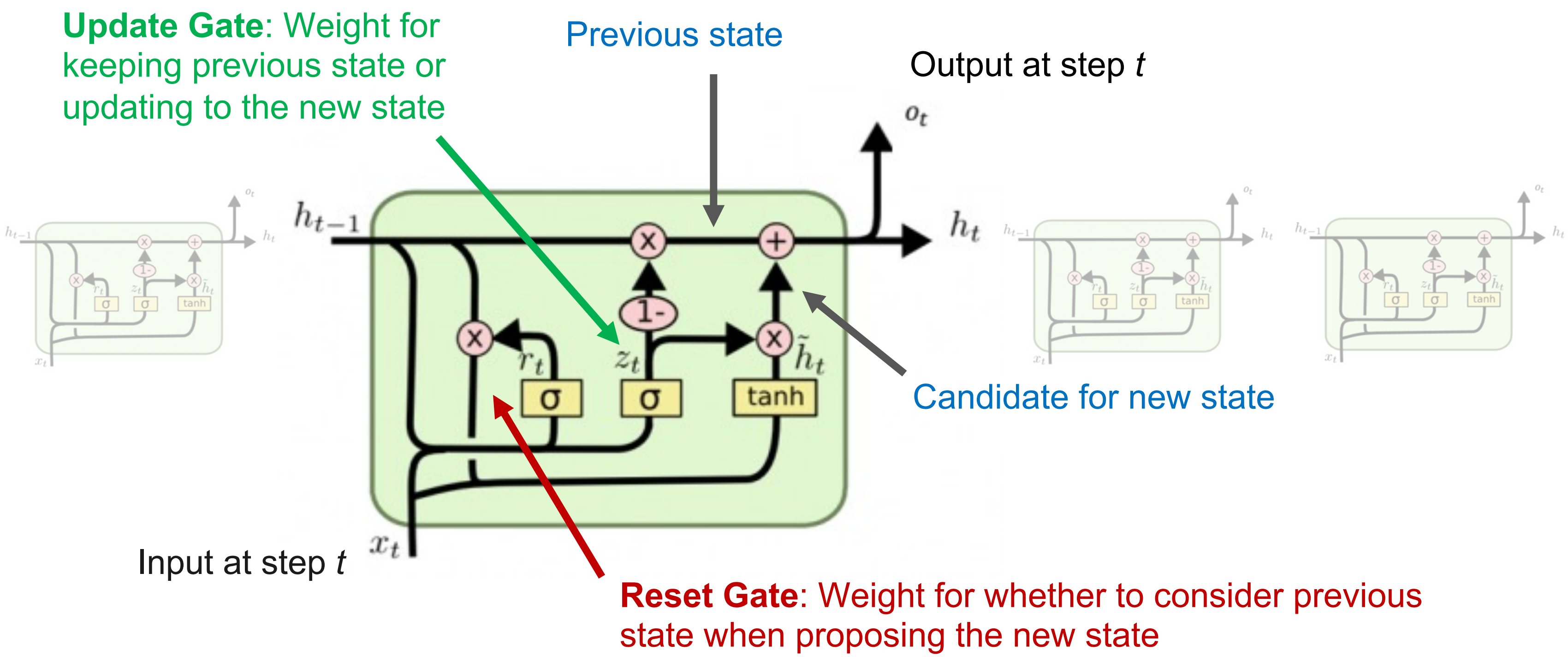
RNN architecture



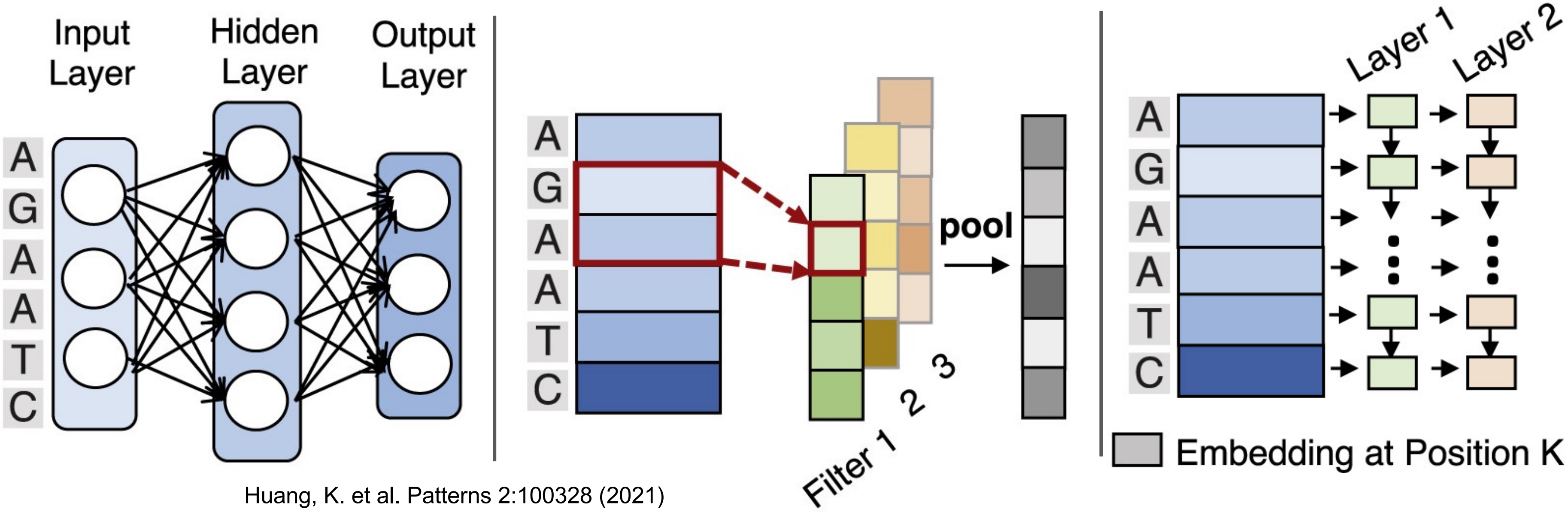
Source: www.linkedin.com/pulse/recurrent-neural-networks-rnn-gated-units-gru-long-short-robin-kalia

- Allow the model to retain / forget information from earlier time points
- Include shortcuts for gradient calculation – similar to ResNet

Gated recurrent unit (GRU)



Picking the right model requires domain knowledge

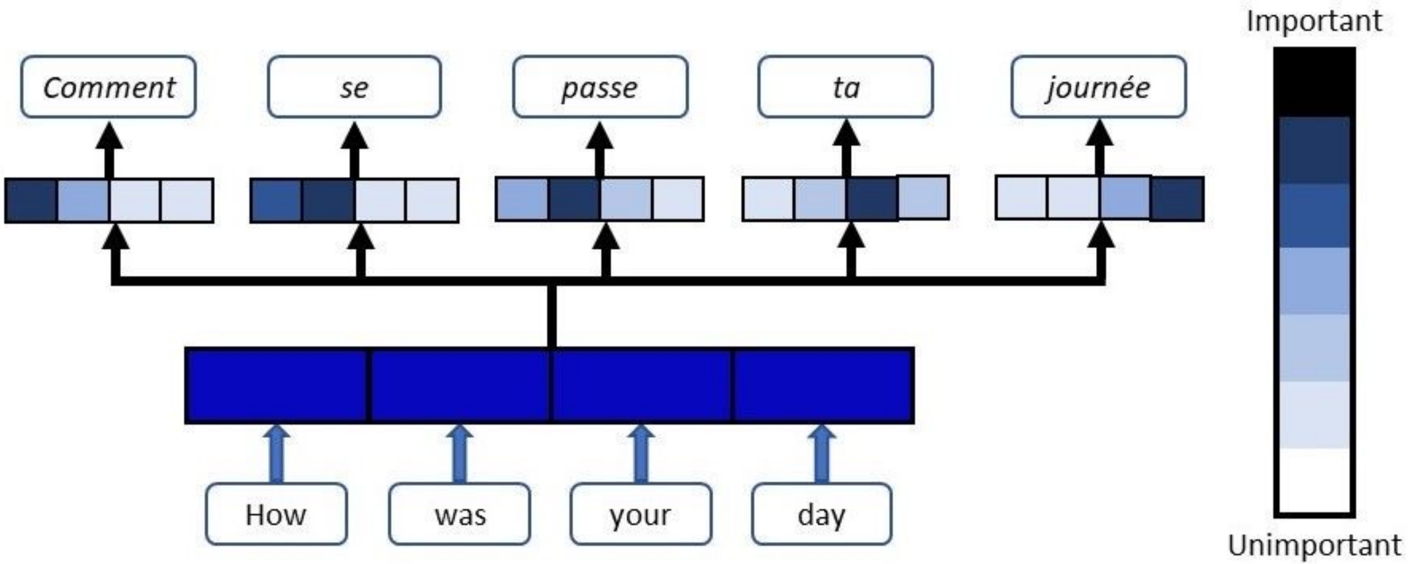


Huang, K. et al. Patterns 2:100328 (2021)

- Choosing the “right” model depends on the interpretation of the task and the underlying mechanisms

Transformer and attention

Attention explicitly model the contribution of each input

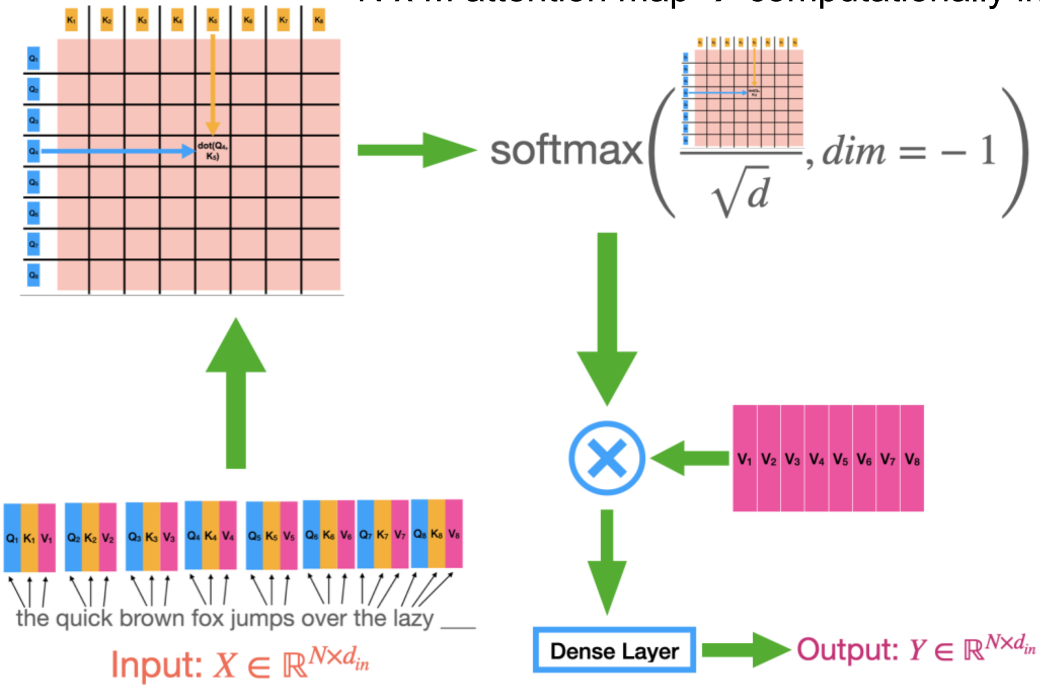


https://blog.floydhub.com/attention-mechanism/

- The model learn to estimate contribution directly during training, not as a post-hoc explainability

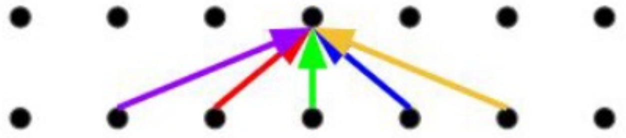
Attention vs regular layers

N x M attention map → computationally intensive



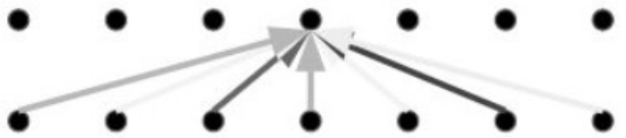
Convolution

Fixed weights regardless of input



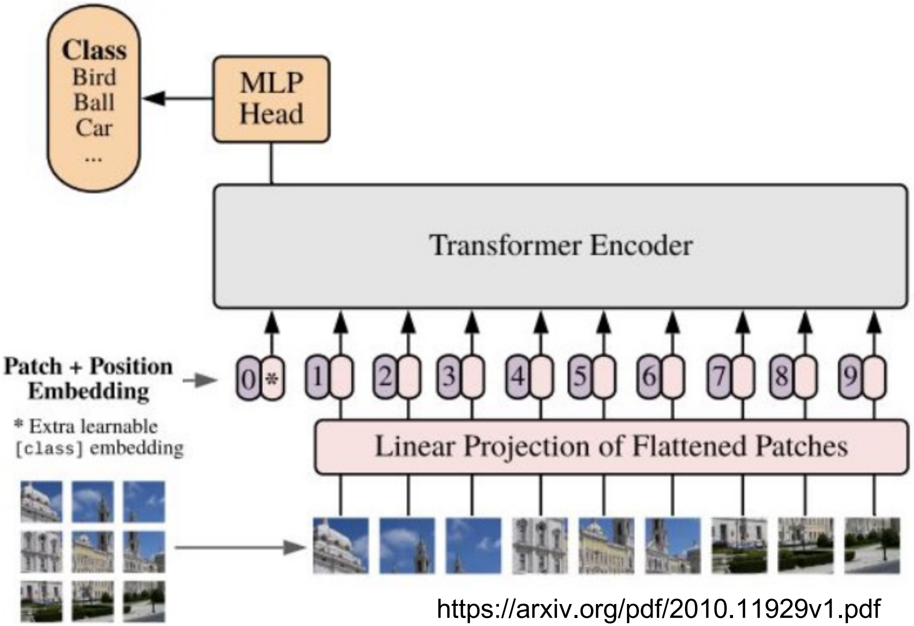
Self-Attention

Adaptive weights depending on input

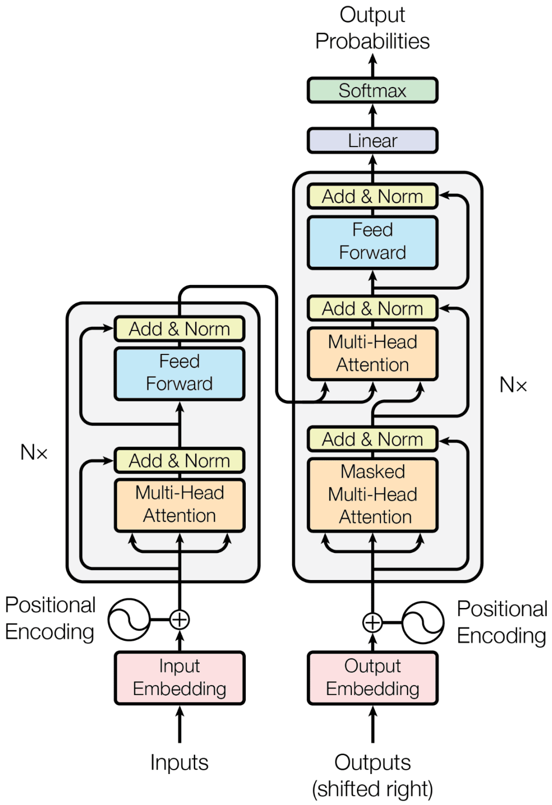


https://learnopencv.com/attention-mechanism-in-transformer-neural-networks/

Transformer architecture

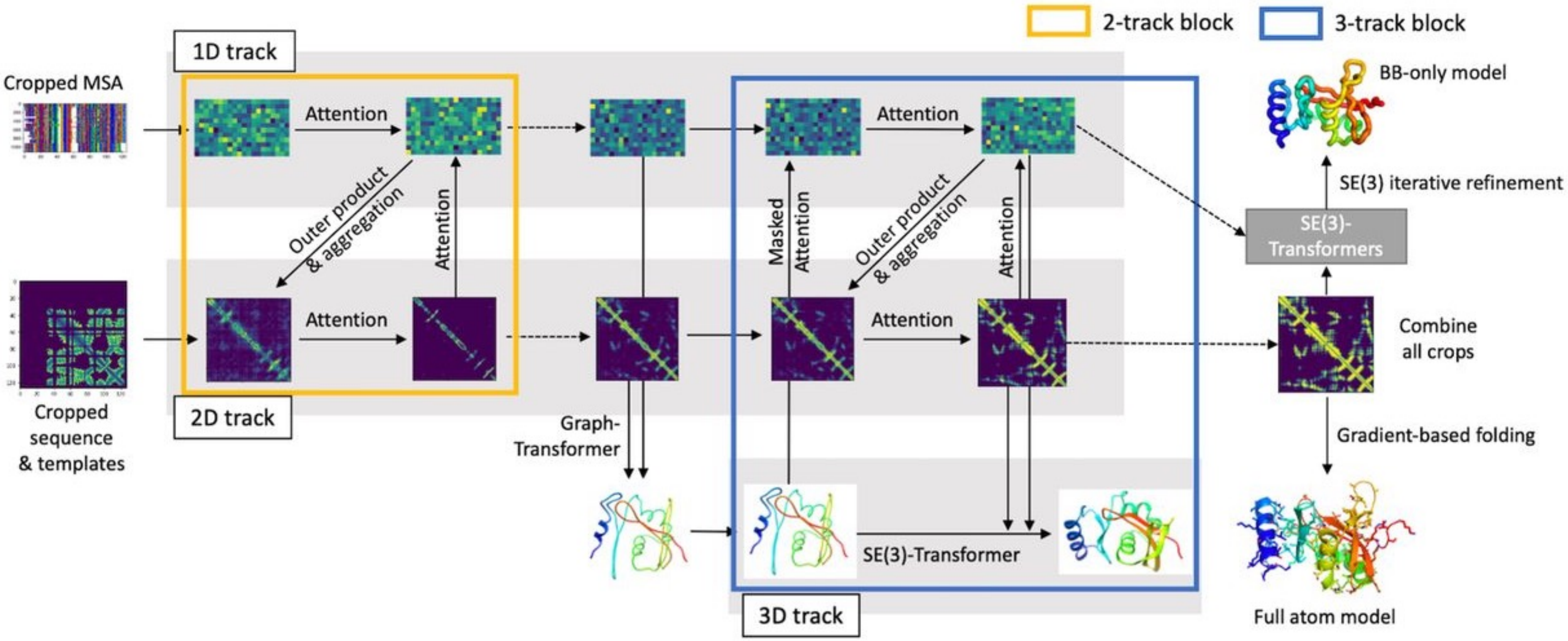


https://arxiv.org/pdf/2010.11929v1.pdf



- Multiple attention layers replace conventional networks

Transformer model for protein folding prediction



- Attention is all you need

Any questions?

See you on March 1st