DL Thinking 1

Konrad Kording and Lyle Ungar

Section 1: Loss Functions

Konrad Kording and Lyle Ungar

Thinking as a data scientist / deep learner

Have more data

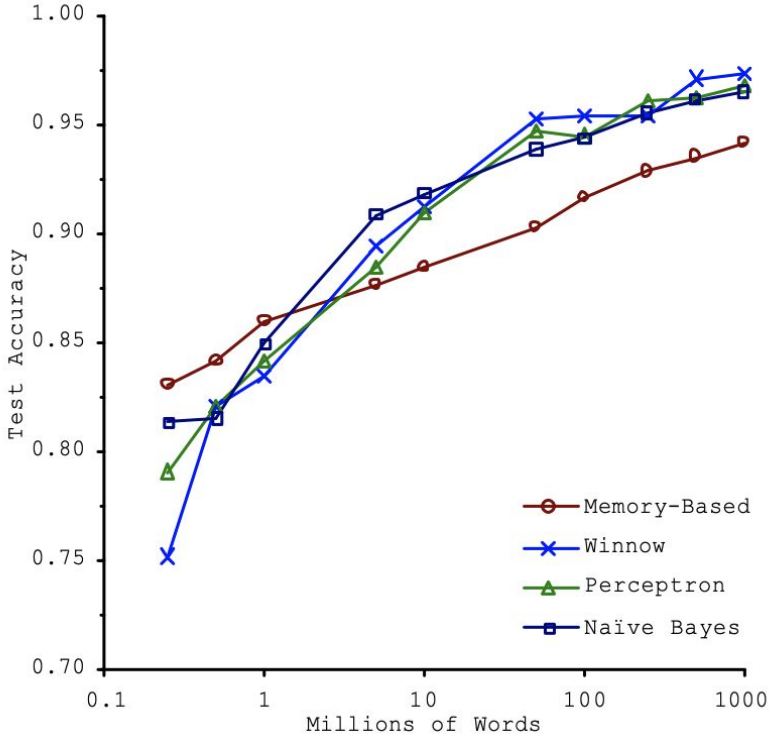
Have better ideas

Image courtesy Brickset/ Flickr

Everyone likes more data

Banko and

Brill, 2001



Ideas do not happen in a vacuum

You know domain

You know experts

You know deep learning

You just need an idea machine!

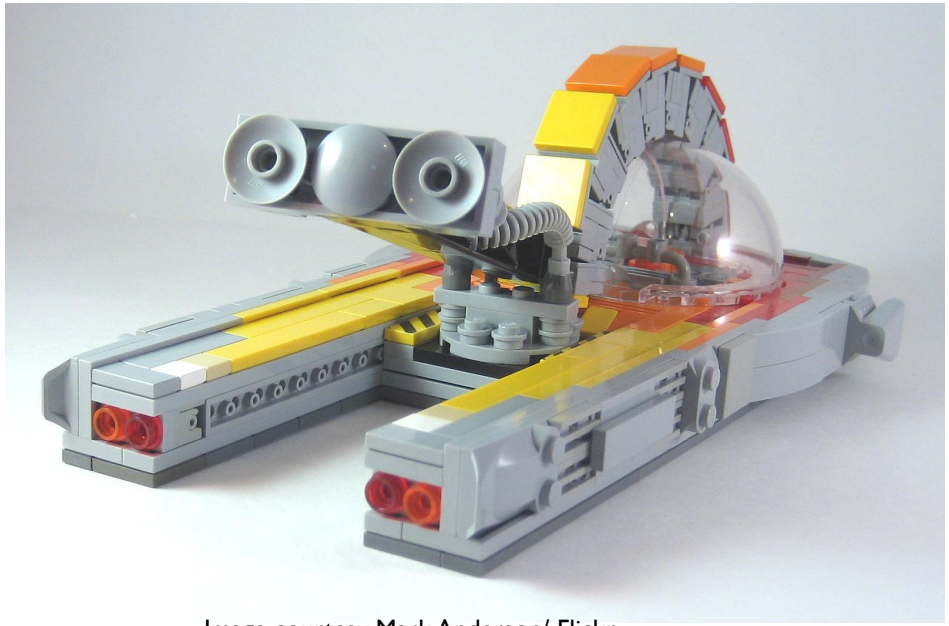


Image courtesy Mark Anderson/ Flickr

Just kidding: but there are some aspects

Clarity about the deliverable

Clarity about domain

Clarity about data

These three mostly define what to do!

What are good ideas in DL?

Image courtesy Brickset/ Flickr

Let us model such a discussion

Image courtesy Orange Xephos/ Flickr

Image courtesy tjparkside/ Flickr

Konrad wears a hat

This is not a slide. We will make the video full screen at this time.

But basically Lyle asks questions. Konrad answers. Learns:

Want to predict what neurons do while motorcycle riding (to fix a brain). Neurons are smooth over 50ms. Movement is smooth over 200ms. Neurons have Poisson noise. Cycle movement is low-d. Highly structured over time.

In practice, this may take a little longer

So you won't do the whole thing

Someone hands you a good ANN

So you just design the cost function

Cost functions

Relevant Information

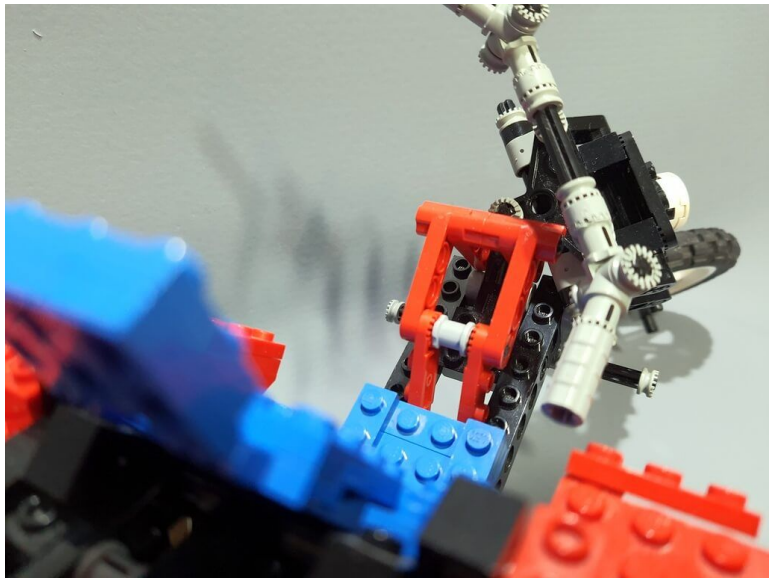


Image courtesy Wikipedia

Cost functions

The most direct Specification of the problem

Clarity=difficult

Here: predict spikes in brain

with subject riding motor cycle

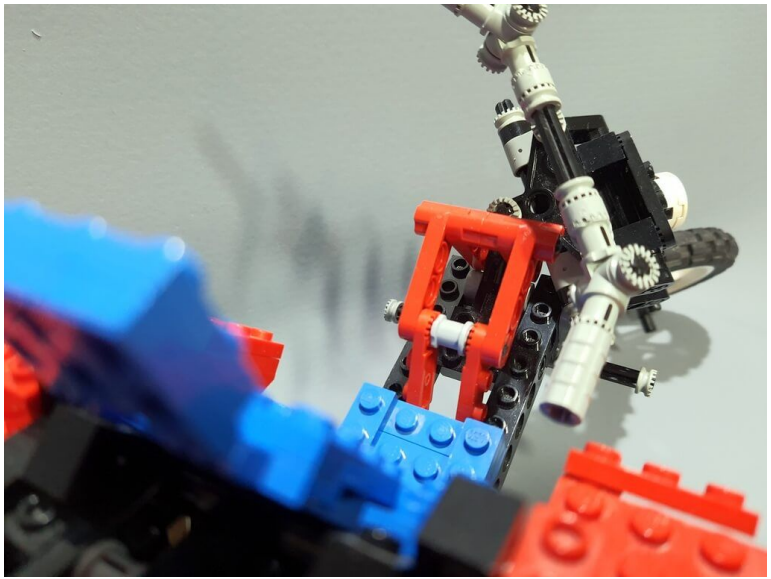
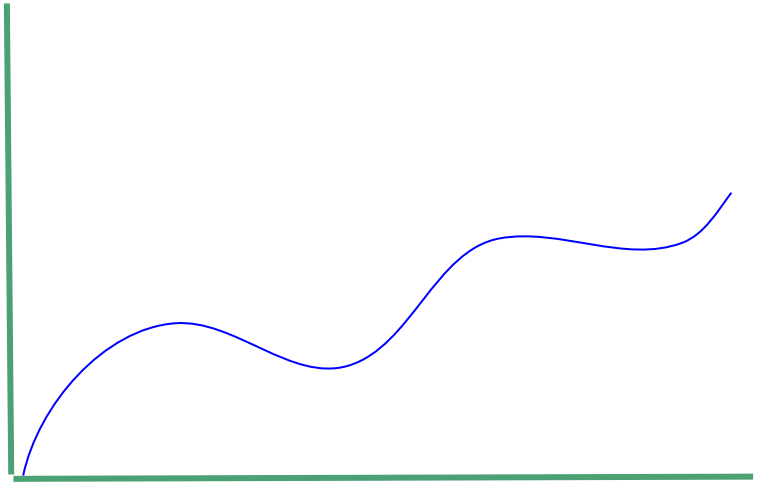


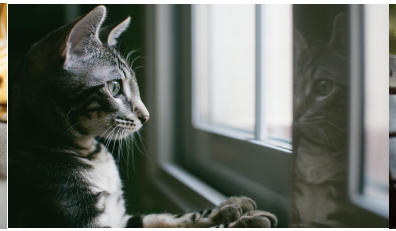
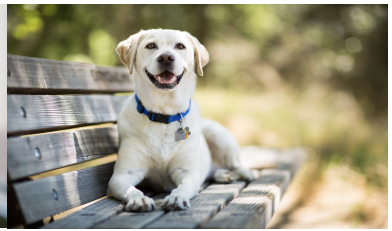
Image courtesy Wikipedia

Regression



Classification







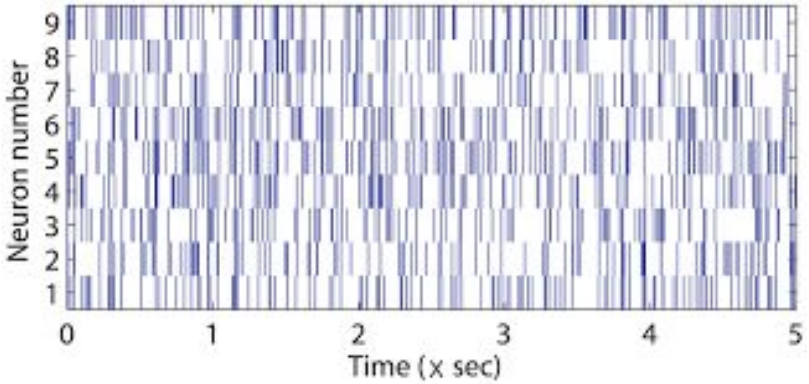
Let us go back to our discussion

Why did Lyle ask his questions?

Which questions relate to cost function?

Lyle asked if the spike data is smooth

Answer was “underlying firing rate is smooth, but every millisecond spikes are random”, inputs are also smooth



Let us think some more about it

Konrad said “every millisecond spiking is random but underlying probability to fire changes slowly”

Means locally every millisecond is independent

So the number of spikes in short interval should be Poisson distributed

$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

So which cost function should we use?

You design the cost function

3 hints and a solution

Hint one: you get time-stamps for the spikes. You will want to do binning. So now you have a bin (say every 50ms) by neuron matrix that contains counts. Now, how do you get a cost function from there?

Hint two: for each bin you can, using machine learning get an estimate of \lambda, the number of spikes expected at that time. For that, ML should get as input the relevant aspects of its inputs at the relevant times (and potentially of the previous times).

Hint three: As bins will be relatively independent, the log-likelihood can be added over all the neurons and time-bins.

Solution: Here is the log-likelihood (\sum_neurons \sum_bins(\lambda_\theta,k_observed))

Advanced: Distributions do not seem quite Poisson. Find a good replacement of Poisson why is this good?

A paper. Is this a neural network?

Published: 06 July 2013

Fast inference in generalized linear models via expected log-likelihoods

Alexandro D. Ramirez & Liam Paninski

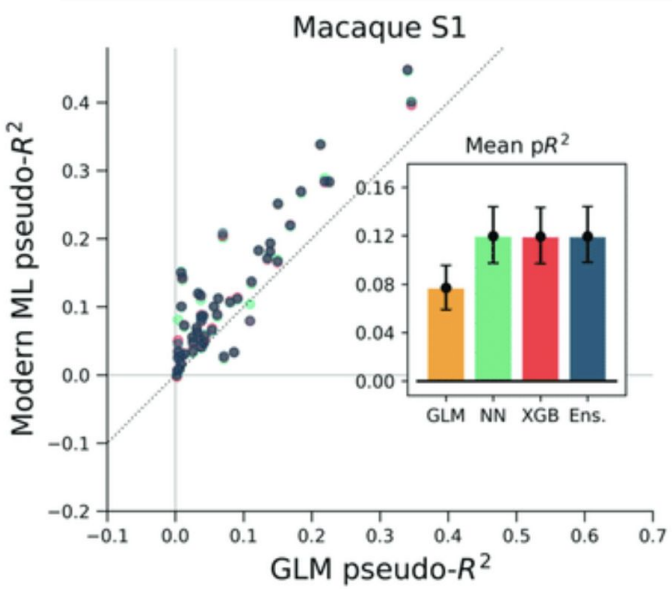
Journal of Computational Neuroscience **36**, 215–234 (2014) | Cite this article

773 Accesses | **15** Citations | Metrics

Abstract

Generalized linear models play an essential role in a wide

Ok. Now we make it big



Machine Learning for Neural Decoding

Joshua I. Glaser, Ari S. Benjamin, Raeed H. Chowdhury, Matthew G. Perich, Lee E. Miller, and Konrad P. Kording

eNeuro 31 July 2020, 7 (4) ENEURO.0506-19.2020; DOI: https://doi.org/10.1523/ENEURO.0506-19.2020

Problem 2: how can an ANN know its uncertainty

There is a big history of focusing machine learning around uncertainty

Bayesian methods

There is a lot of blowback now from DL people

Lyle wears a hat

This is not a slide. We will make the video full screen at this time.

But basically Konrad asks questions. Lyle answers. Learns:

We do regression. We want to know how uncertain we are. Why not just take MSE? Some items are easier vs harder. Why does it help to know uncertainty? Because confidence matters for actions. Examples: drug effects, Images of numbers, etc, Chemical spectrum, What form of uncertainty may we expect? Gaussian

What do we want?

Every estimate should have an associated uncertainty value

Lets talk:

Cost functions

Relevant Information

Image courtesy Wikipedia / spring of hope

Cost functions

Predict uncertainty from stimuli

Image courtesy Wikipedia / spring of hope

Let us go back to our discussion

Why did Konrad ask his questions?

Which questions relate to cost function?

You want to know uncertainty of your ANN

We estimate the location of atoms in a Chemical molecule

Based on all kinds of inputs

But now we want to have mean and variance of our estimates

Here we have a Gaussian

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

What do we want

For every stimulus

- Estimate
- Uncertainty of estimate

So which cost function should we use?

You design the cost function

TA guide

Hint 1: look at the Gaussian. Where is there the estimate? Where is there the uncertainty?

Hint 2: We generally need log-probabilities as they are additive over samples. write out the log likelihood function of the Gaussian

Solutions: Maximize sum log(gaussians (data-model))

Advanced: log sigma can obviously fail for negative sigma. What now?

Example: predict NMR

Rapid prediction of NMR spectral properties with quantified uncertainty

Eric Jonas & Stefan Kuhn

Journal of Cheminformatics **11**, Article number: 50 (2019) | Cite this article

And invert this

Deep imitation learning for molecular inverse problems

Eric Jonas
Department of Computer Science
University of Chicago
ericj@uchicago.edu

Problem 3: embed faces well

Konrad wears a hat

This is not a slide. We will make the video full screen at this time.

So, I want to recognize faces. One photo. Always recognize you. Because I am kinda face blind. I want 1 shot. Idea. Project into face space. Close faces are likely to be same person. Lyle: but what about different faces. Should be far. But what does far mean? Dunno. Just standard 2 metric.

What do we want?

Good embedding

Lets talk:

Cost functions

Relevant Information

You want to embed faces so you can face recognize based on distance

Why?

How would you do it?

Image courtesy Wikipedia / faces (band)

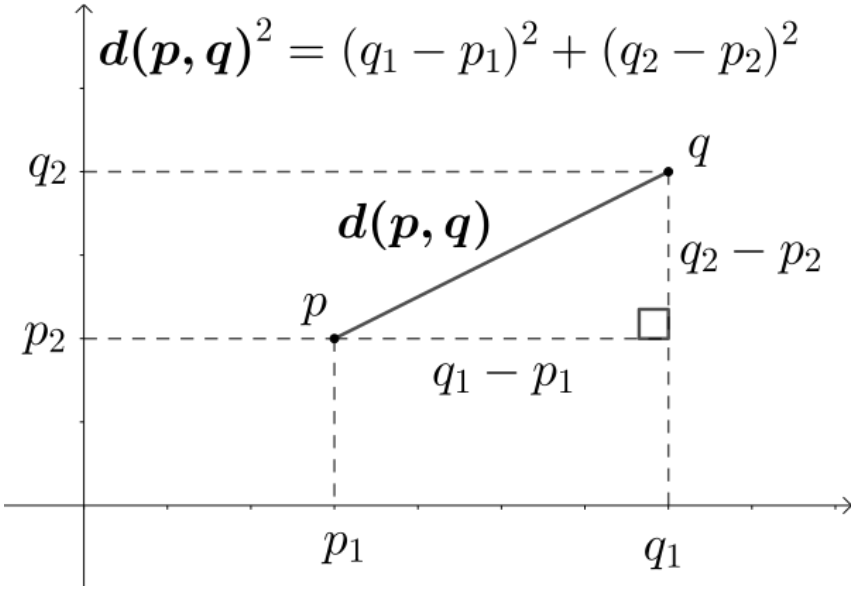
Have face dataset

Image courtesy yale face dataset

Embedding idea

$$y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Euclidean Metric



What do we want

Cost function that measures quality of embedding

So which cost function should we use?

You design the cost function

TA guide

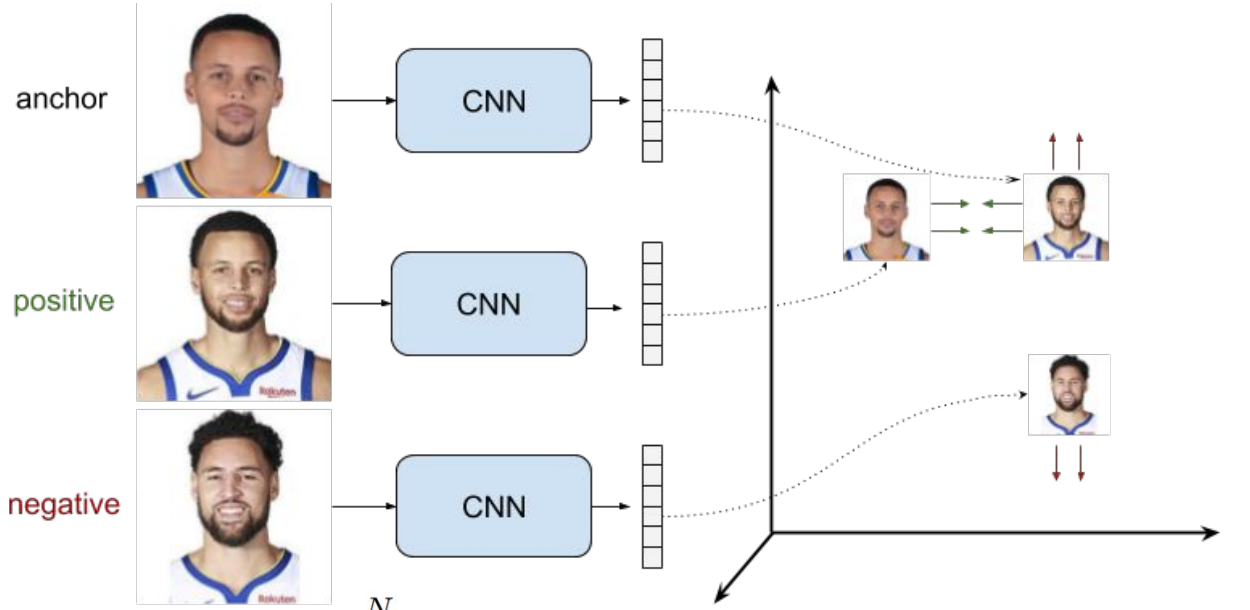
Hint 1: how do we want to deal with same faces? What about different faces?

Hint 2: Different faces to be far, Same faces to be near. Can we phrase this with 3 faces?

Solution: Triplet loss (you load 3 images. 2 identical, 2 different, cost is distance between same - distance between different.

Advanced: Discuss how we could improve over this (e.g. by matching face type), discuss race/ gender

Triplet loss



$$\sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]$$

Large Scale Online Learning of Image Similarity Through Ranking

Gal Chechik GAL@GOOGLE.COM
Google
1600 Amphitheatre Parkway
Mountain View CA, 94043

Varun Sharma* VASHARMA@GOOGLE.COM
Google, RMZ Infinity
Old Madras Road, Bengalooru
Karnataka 560016, India

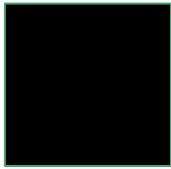
Uri Shalit*† URI.SHALIT@MAIL.HUJI.AC.IL
The Gonda Brain Research Center
Bar Ilan University
52900, Israel

Samy Bengio BENGIO@GOOGLE.COM
Google
1600 Amphitheatre Parkway
Mountain View CA, 94043

What would you do if the world was adversarial?

How would you deal with masks?

How would you think about the ethics of this?



Bonus

TA guide (only for TAs that are comfortable with this)

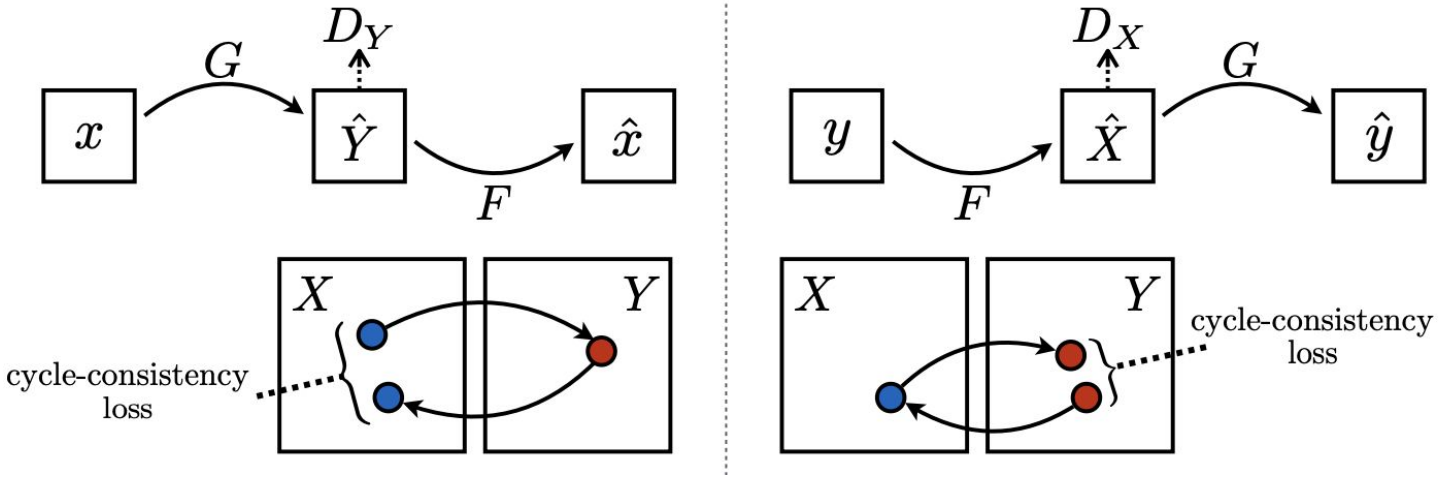
This time the students ask you. You want to build something that makes winter photos into summer photos. And vice versa. The idea is that it is like generating images for winter and summer. But what could we use extra for this? Well, a winter photo made into a summer photo and back should be similar to the original image. This is the basis of cycle gan

You want to build a system that converts summer pictures to winter pictures

Because skiing

What would be a good setting?

Cycle GAN loss



Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Jun-Yan Zhu* Taesung Park* Phillip Isola Alexei A. Efros
Berkeley AI Research (BAIR) laboratory, UC Berkeley