ML4CC: Lecture 4

Sit with your discussion groups (same as last time)!

Assignments reminder

Keep doing your weekly PMIRO+Q

Your first coding assignment was due at 8am.

Your second coding assignment will be posted after class and is due Feb 29 before the start of class.

Recap of previous paper

P: Need to be able to map building damage after a disaster

M: Train a convolutional neural network to classify building damage using an existing dataset of satellite imagery

I: Vary the loss function and the type of inputs provided; also applied a “saliency” method

R: Ordinal cross entropy loss with all three inputs (pre/post image with disaster type) performs best

O: The dataset was filtered to not include small buildings and was subsampled to be balanced across classes

Climate Change in the News

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ | Subscribe

nature > news > article

NEWS | 09 February 2024 | Clarification 12 February 2024

Climatologist Michael Mann wins defamation case: what it means for scientists

Jury awards Mann more than US$1 million – raising hopes for scientists who are attacked politically because of their work.

By Jeff Tollefson

The case stems from a 2012 blog post published by the Competitive Enterprise Institute (CEI), a libertarian think-tank in Washington DC. In it, policy analyst Rand Simberg compared Mann, then at Pennsylvania State University in State College, to Jerry Sandusky, a former football coach at the same university who was convicted of sexually assaulting children, saying that "instead of molesting children, ***he has molested and tortured data in the service of politicized science*** that could have dire economic consequences for the nation and planet". Author Mark Steyn subsequently reproduced Simberg's comparison as he ***accused Mann of fraud*** in a blog published by the conservative magazine *National Review*. In the same year, **Mann sued both Simberg and Steyn**, as well as the CEI and the *National Review*, for libel, without asking for damages.

As a public figure, Mann and his attorneys ***had to prove not only that the defendants published false statements, but also that they acted with malice***. "It is not easy to prove defamation against a public figure," says Lauren Kurtz, executive director of the Climate Science Legal Defense Fund, an organization in New York City that was formed in 2011 to advocate for Mann and other scientists who were being targeted and harassed by climate-change sceptics.

Paper 3 Discussion

Poster at NeurIPS - Tackling Climate Change with Machine Learning workshop 2023

Detailed Glacier Area Change Analysis in the European Alps with Deep Learning

Codruț-Andrei Diaconu[1,2] **Jonathan L. Bamber**[2,3]
codrut-andrei.diaconu@dlr.de j.bamber@tum.de
[1]German Aerospace Center (DLR) [2]Technical University of Munich [3]University of Bristol

Abstract

Glacier retreat is a key indicator of climate change and requires regular updates of the glacier area. Recently, the release of a new inventory for the European Alps showed that glaciers continued to retreat at about 1.3% a^{-1} from 2003 to 2015. The outlines were produced by manually correcting the results of a semi-automatic method applied to Sentinel-2 imagery. In this work we develop a fully-automatic pipeline based on Deep Learning to investigate the evolution of the glaciers in the Alps from 2015 to present (2023). After outlier filtering, we provide individual estimates for around 1300 glaciers, representing 87% of the glacierized area. Regionally we estimate an area loss of -1.8% a^{-1}, with large variations between glaciers. Code and data are available at https://github.com/dcodrut/glacier_mapping_alps_tccml.

Attendance

Select one person from the group to go to this Google Doc and write down the names of all people present in the group (remember to mark who took attendance!)

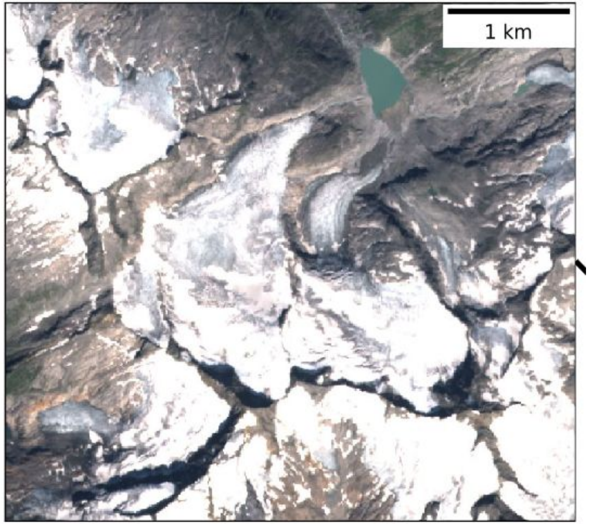
https://docs.google.com/document/d/1PKhw9E2lJpAnFrFO88DOc2rscZFVlcIv47QNa5h1sGs/edit?usp=sharing (link is in Brightspace under Syllabus content)

Discussion Question 1

Explain what the authors say they are doing here, why they are doing it, and whether or not you think it is a good idea.

Given that the resolution of Sentinel-2 data is 10m, we decided to use only the glaciers with an area larger than 0.1 km^2. Although this reduces the number of glaciers sampled (1646, *i.e.* about 37%), the percentage of glacierized area covered is close to 95%. To facilitate further analyses, we additionally

Removing small glaciers



Getting rid of glaciers that are fewer than 10 pixels removes the majority of individual glaciers.

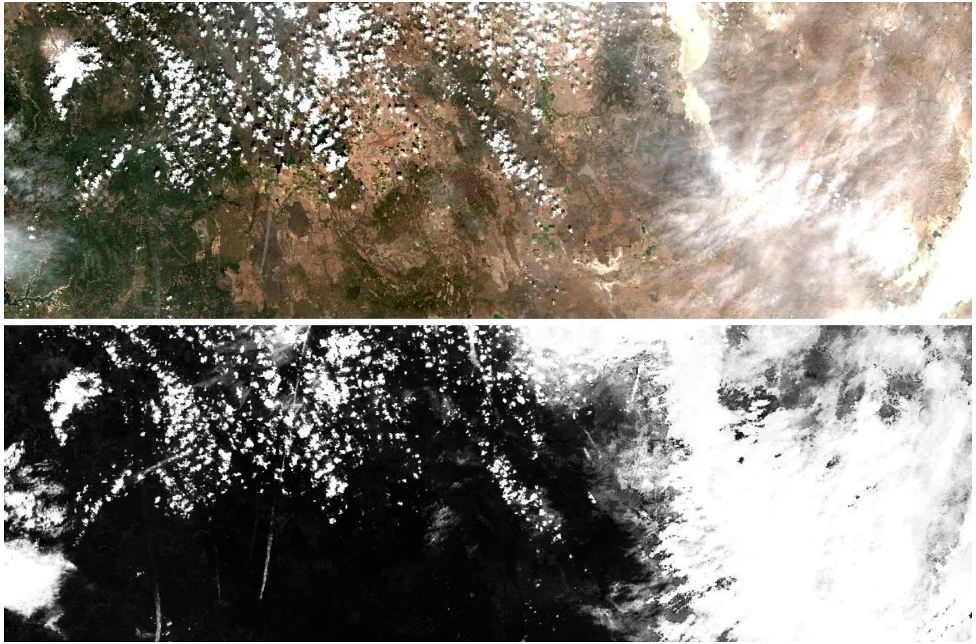
But, because there are many large glaciers, it does not significantly impact the total amount of glacier cover included in the data (~95%), which is arguably the more important measure here.

Discussion Question 2

Explain what the authors say they are doing here, why they are doing it, and whether or not you think it is a good idea.

> A common criterion used to download optical data is by choosing the tile (*i.e.* 100km x 100km for Sentinel-2) with the smallest percentage of cloud coverage for each region of interest. If we follow this strategy and then compute the average cloud coverage per glacier (using the inventory outlines), we obtain an average of 4%. However, if rather than restricting to one single tile we instead use the least cloudy five tiles (centered on 01.09.2023 $\pm$ 15 days) and then choose the best for each glacier individually, we significantly reduce the cloud coverage to 0.1%. This is lower than the average in

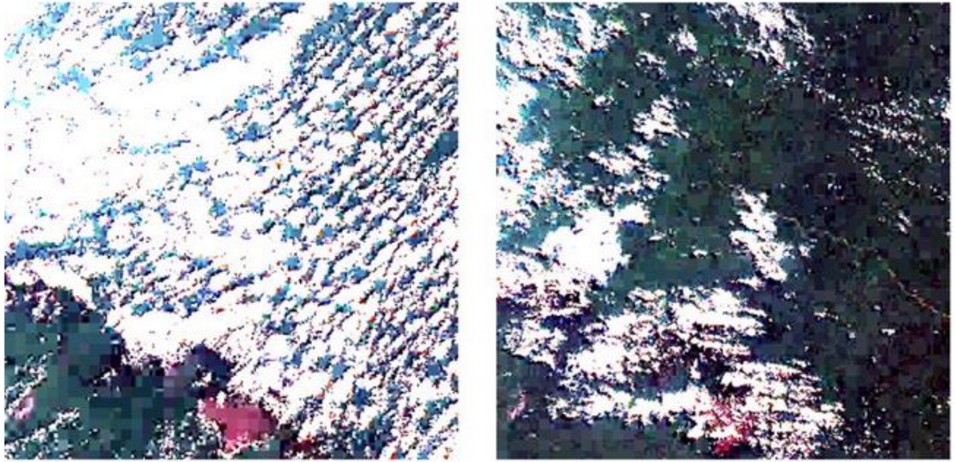
Sentinel-2 Cloud Coverage



Sentinel-2 satellite imagery (top) comes with cloud probability masks (bottom).

This lets you choose which image date you want to use based on which day has the lowest cloud coverage

Sentinel-2 Cloud Coverage



(d) 2017 01 19 (e) 2017 08 07

Kempineers and Soille

It is possible (especially in 100km by 100km images) that the least cloudy image overall might not be the best for every individual pixel.

Therefore they pick the best image for each glacier individually (out of the top 5 least cloudy)

Discussion Question 3

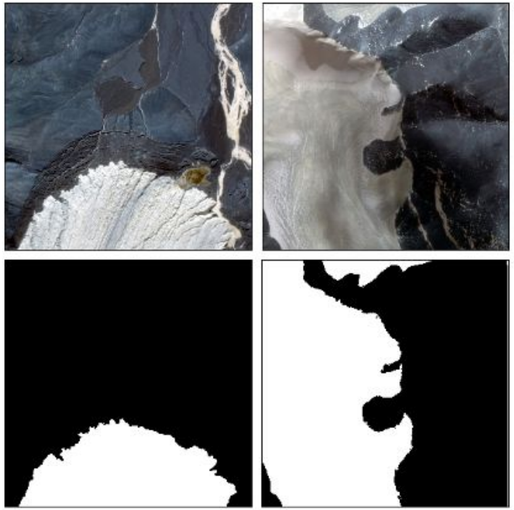
What loss function was used to train the model? Why is this appropriate for the problem?

Loss function

Binary cross entropy loss

$$H(P^*|\,P) = -\sum_i \underbrace{P^*(i)}_{\text{TRUE CLASS DISTIRBUTION}} \log \underbrace{P(i)}_{\text{PREDICTED CLASS DISTIRBUTION}}$$

Binary segmentation (e.g. glacier vs not glacier) can be understood as a binary classification task for each pixel



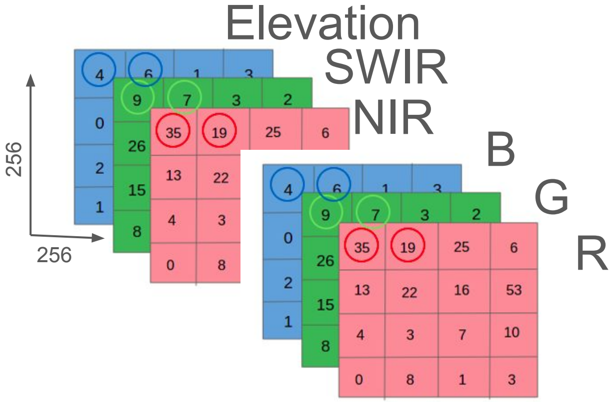
Chu et al

Discussion Question 4

What are the dimensions of the input to the model?

Input

256x256x6



performing in [8], with a relatively smaller model size compared to the other methods evaluated. We extend the input from three to six channels, to accommodate the following inputs:

- five Sentinel-2 bands: blue (B2), green (B3), red (B4), NIR (B8) and SWIR (B12), which we found the most informative;
- surface elevation, obtained from NASADEM[17] (30m resolution) and processed using the Open Global Glacier Model[18]. The surface elevation should help for debris-cover on glaciers [10], Central Europe being one of the regions with the highest percentage of debris cover[19].

Model training. We train the model to predict the probability that each pixel is glacier or not, using patches of 256x256 pixels (*i.e.* 2.56x2.56 km). Given that we apply the model only on glacierized

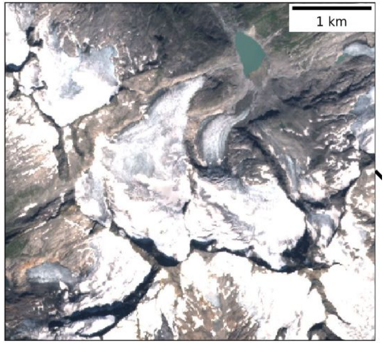
Larger images are cut into smaller ones

Discussion Question 5

What was done to help balance the data? How does it help? Do you think it was a good idea or not?

Centered images

> patches of 256x256 pixels (*i.e.* 2.56x2.56 km). Given that we apply the model only on glacierized regions, we sample patches only if the center is on the glacier, which also helps in balancing the two classes. This implies that the model sees only the glaciers and a maximum buffer of 1.28km around,



Sampling mountain images at random may lead to a lot of “not glacier” pixels. Centering images around glaciers ensures a similar mix of glacier and non-glacier.

Sometimes such centering would be “cheating” because at run time you may not know where the glaciers are. But the goal here is to track the size of glaciers that were already inventoried, so it is acceptable to use that information.

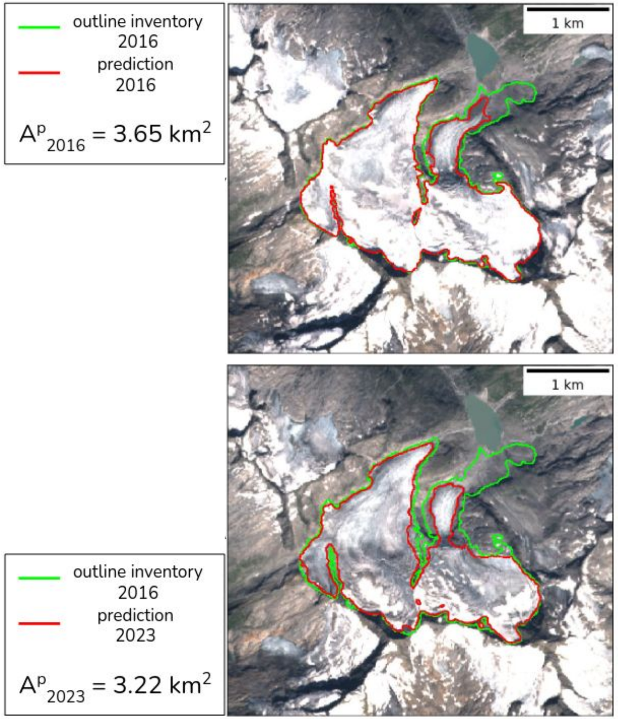
Discussion Question 6

Why will the errors “cancel out” here?

Area (change) estimation. Given the significant volume loss observed over the 2000-2019 period[5], with a mean elevation change of -1.02 $\pm$ 0.21 m a^{-1}, we can assume that glaciers in this region do not grow over the 2015-2023 period. This allows us to extract the changes in the areas by applying the model for each glacier but only for the pixels within the inventory outlines, thus excluding the predictions outside these. However, we do not use the areas from the inventory as the reference value but the predicted ones such that, if the model makes systematic errors, they will cancel out, as in the case illustrated in Figure 1. Therefore, for each glacier, we calculate the area change per year as

The model should underestimate in the same way across time

Here the model (red) underestimates the size of the glacier compared to how it was labeled in the dataset (green) in both 2016. Comparing the model's output in 2023 to the labeled data from 2016 would overestimate the amount of glacier melt. But using the model's estimates at both time points does not.



Discussion Question 7

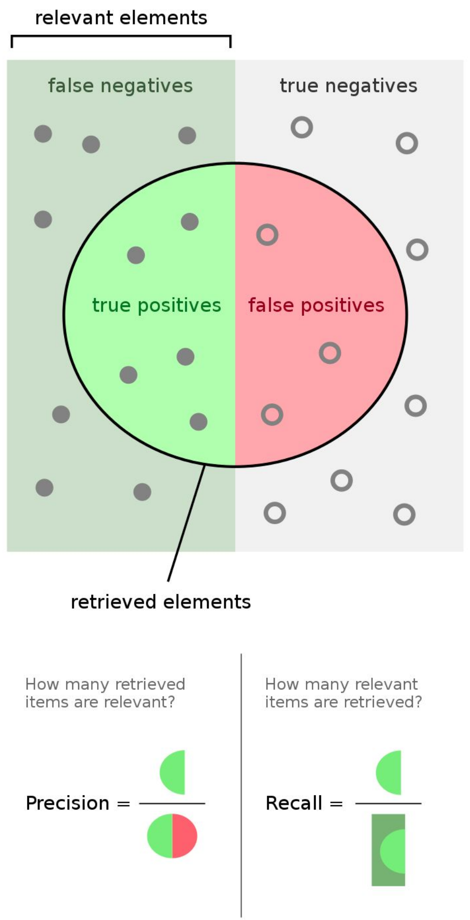
Which does the model do better at: labeling glacier pixels as glaciers or not labeling non-glacier pixels as glaciers?

Precision and Recall

Table 1: Performance metrics for each of the five testing CV folds.

subregion	#patches	#glaciers	Accuracy	IOU	Precision	Recall	F1
r_1	1855	349	0.953	0.794	0.875	0.896	0.878
r_2	1321	234	0.955	0.862	0.924	0.926	0.923
r_3	1084	184	0.960	0.879	0.931	0.937	0.933
r_4	2146	406	0.964	0.836	0.916	0.903	0.905
r_5	2301	437	0.951	0.769	0.951	0.796	0.857
		$\mu \pm \sigma$:	0.96 ± 0.01	0.83 ± 0.05	0.92 ± 0.03	0.89 ± 0.06	0.90 ± 0.03

Precision is slightly higher than recall, so the model is a slightly better at not labeling non-glaciers as glaciers (though both numbers are high).



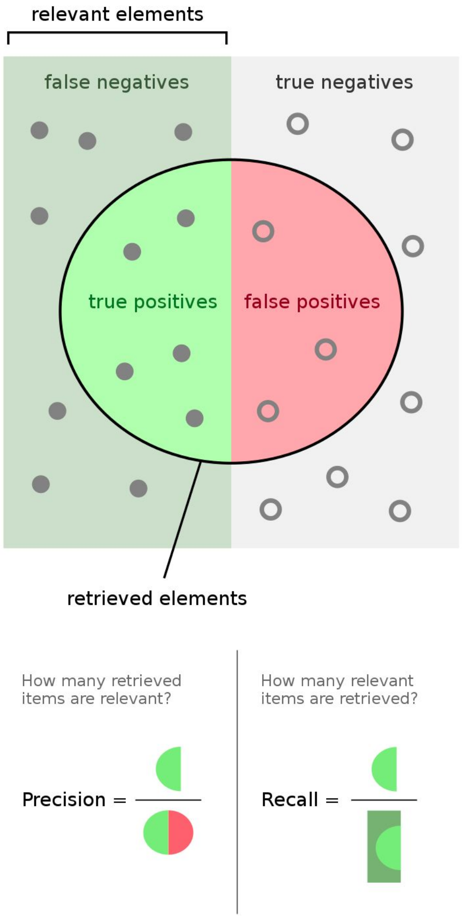
Precision and Recall

Follow-up Q: what are the rows in this table?

Table 1: Performance metrics for each of the five testing CV folds.

subregion	#patches	#glaciers	Accuracy	IOU	Precision	Recall	F1
r_1	1855	349	0.953	0.794	0.875	0.896	0.878
r_2	1321	234	0.955	0.862	0.924	0.926	0.923
r_3	1084	184	0.960	0.879	0.931	0.937	0.933
r_4	2146	406	0.964	0.836	0.916	0.903	0.905
r_5	2301	437	0.951	0.769	0.951	0.796	0.857
		$\mu \pm \sigma$:	0.96 ± 0.01	0.83 ± 0.05	$\boxed{0.92 \pm 0.03}$	$\boxed{0.89 \pm 0.06}$	0.90 ± 0.03

Precision is slightly higher than recall, so the model is a slightly better at not labeling non-glaciers as glaciers (though both numbers are high).



Precision and Recall

Follow-up Q: what are the rows in this table?

Table 1: Performance metrics for each of the five testing CV folds.

subregion	#patches	#glaciers	Accuracy	IOU	Precision	Recall	F1
r_1	1855	349	0.953	0.794	0.875	0.896	0.878
r_2	1321	234	0.955	0.862	0.924	0.926	0.923
r_3	1084	184	0.960	0.879	0.931	0.937	0.933
r_4	2146	406	0.964	0.836	0.916	0.903	0.905
r_5	2301	437	0.951	0.769	0.951	0.796	0.857
		$\mu \pm \sigma$:	0.96 ± 0.01	0.83 ± 0.05	0.92 ± 0.03	0.89 ± 0.06	0.90 ± 0.03

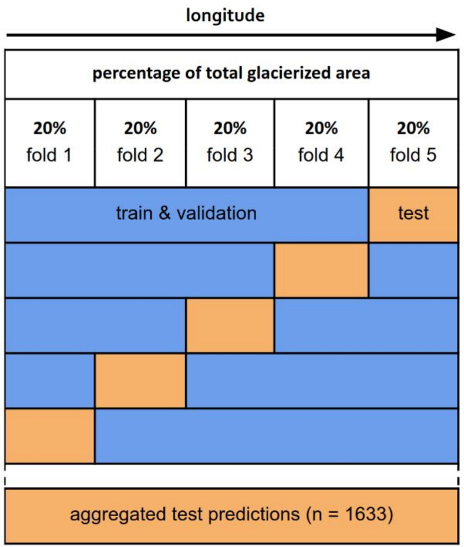


Figure A1: Cross-validation scheme with a geographic split

Discussion Question 8

Share what questions you wrote in your PMIRO+Q and decide as a group what you'd like to ask.

Update your PMIRO+Q

Submit a second file to the Brightspace assignment (don’t overwrite the original):

It should:

Update your PMIRO as needed

Answer your own Q

You can be talking with your group during this!

15 min break

Lecture

Climate content: weather and climate simulation models

Machine learning content: generative models, evaluation metrics, self-supervised learning

Shared Socio-Economic Pathways (SSPs)

- 5 narrative scenarios, representing different approaches and resulting challenges
- SSPs include social, economic, and governmental forces and challenges
- Provide a framework for making predictions of possible futures

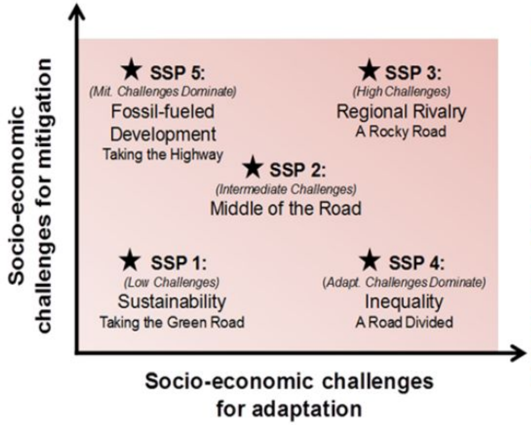
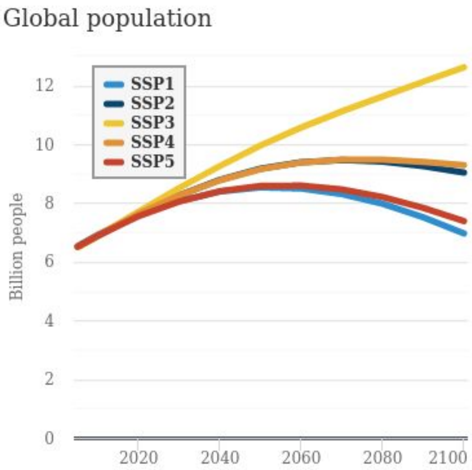
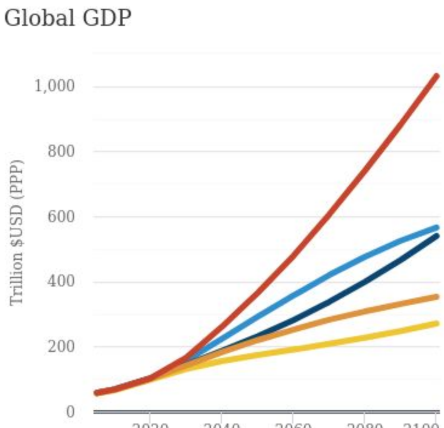


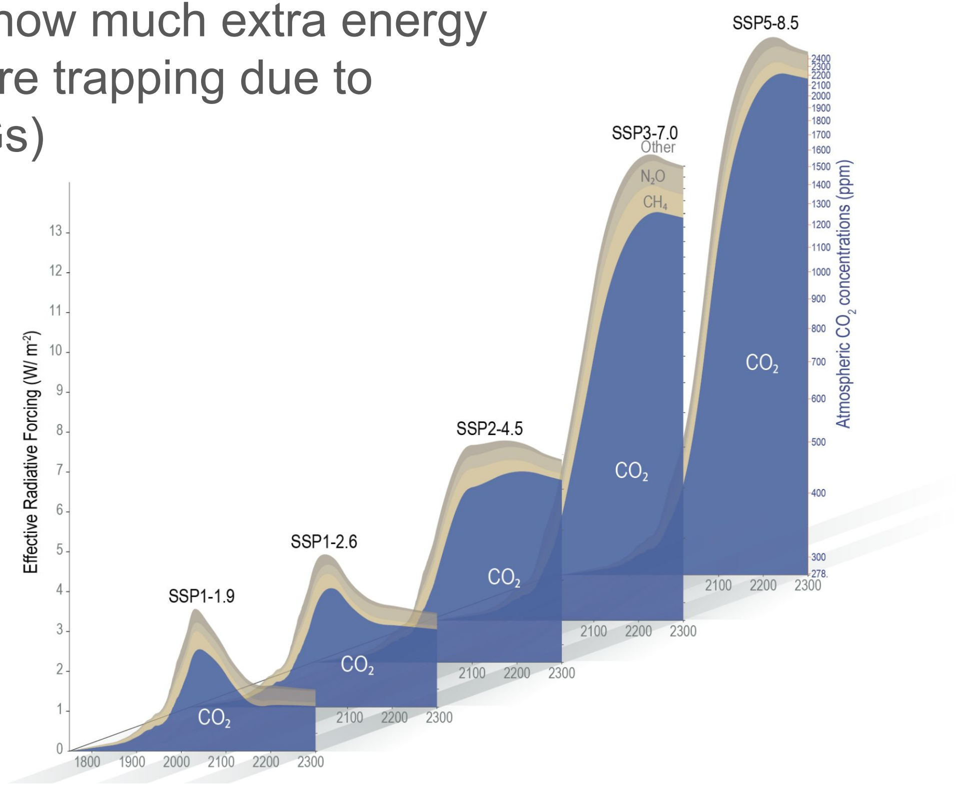
Fig. 1 Overview of SSPs
(Narratives in O'Neill et al., 2016, Glob Env Change, online first)
SSP1: low challenges for mitigation (resource efficiency) and adaptation (rapid development)
SSP3: high challenges for mitigation (regionalized energy / land policies) and adaptation (slow development)
SSP4: low challenges for mitigation (global high tech economy), high for adapt. (regional low tech economies)
SSP5: high challenges for mitigation (resource / fossil fuel intensive) and low for adapt. (rapid development)

SSPs as possible futures

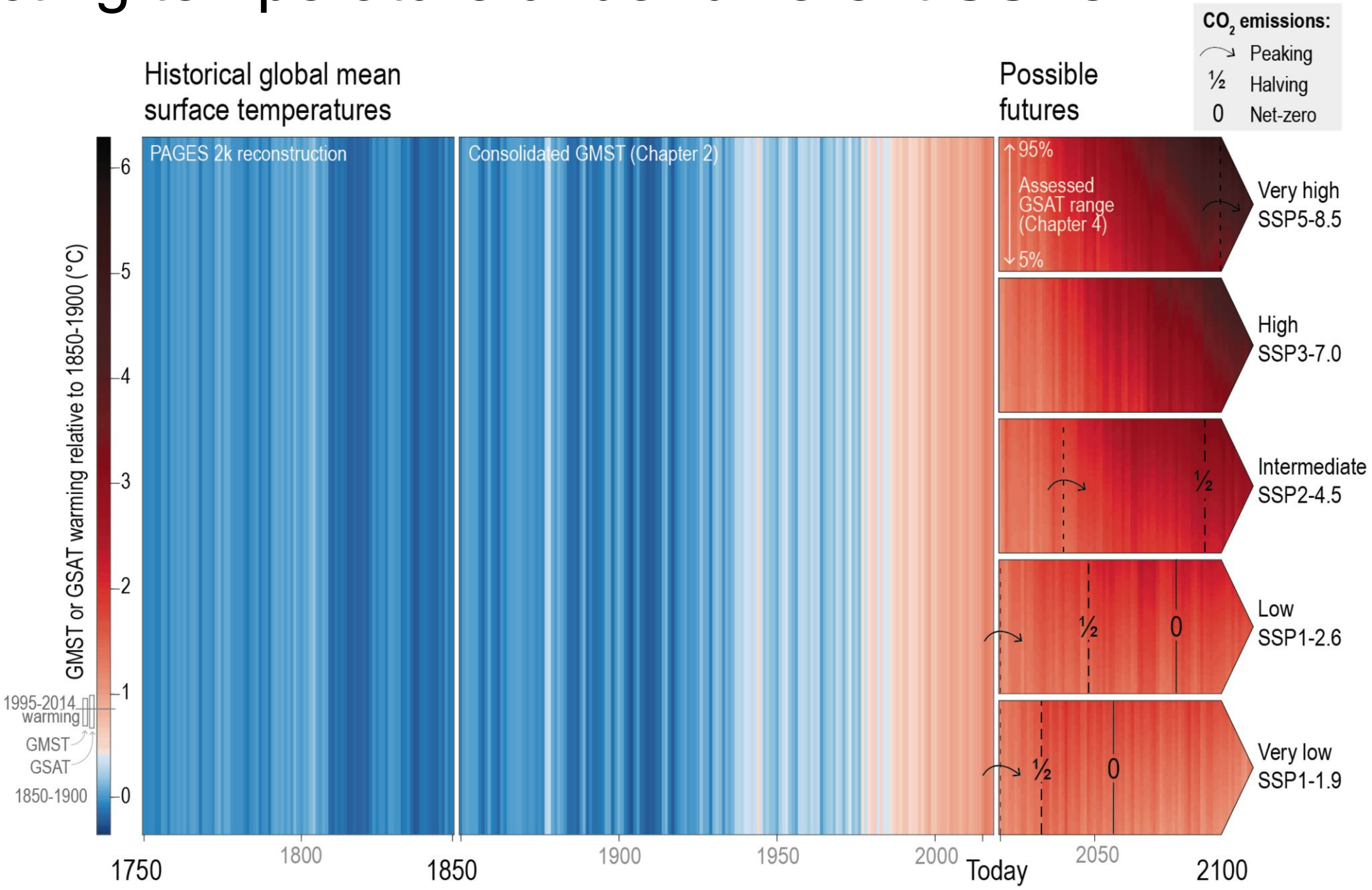




Also labeled according to amount of radiative forcing (i.e. how much extra energy we are trapping due to GHGs)

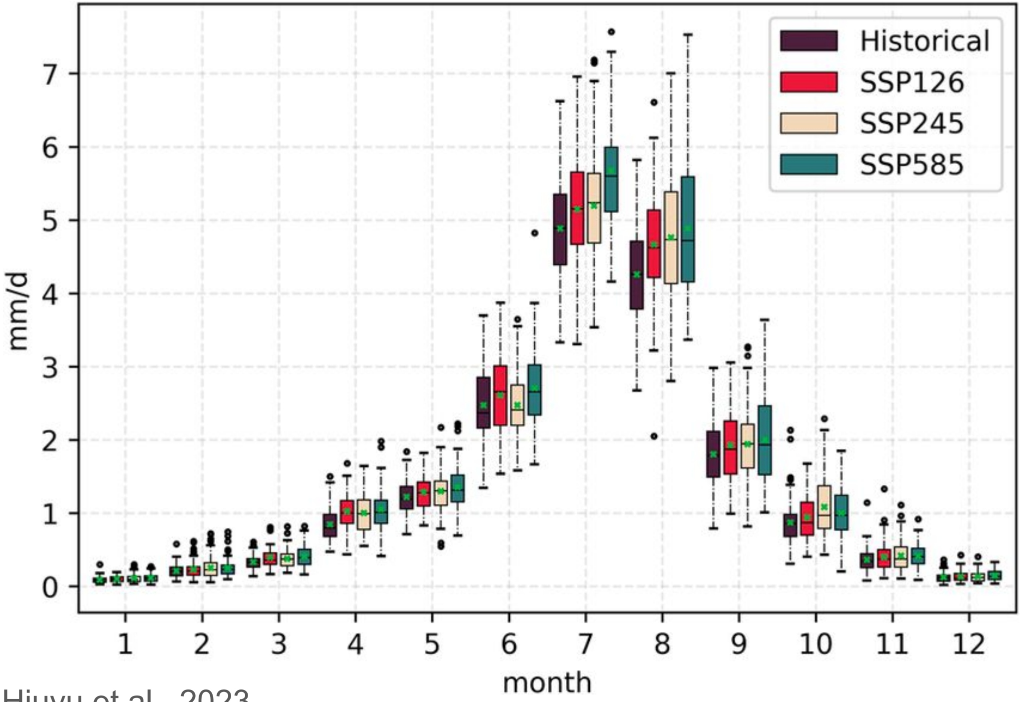


Predicting temperature under different SSPs



Predicting precipitation under different SSPs

Monthly precipitation



Hiuyu et al., 2023

How do we make these predictions?

Climate simulation models!

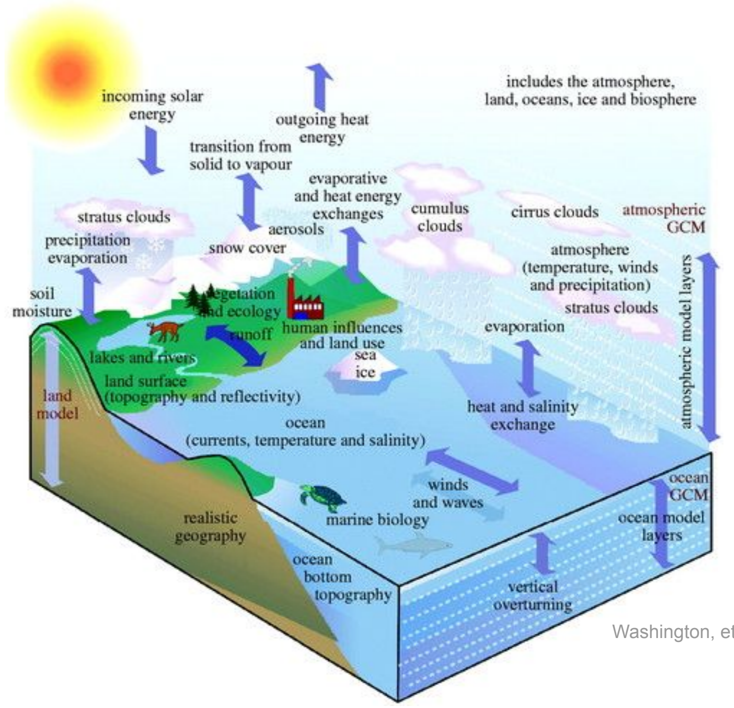
A global climate model (GCM) is a complex mathematical representation of the major climate components and their interactions. The main climate system components treated in a climate model are:

The atmospheric component, which simulates clouds and aerosols, and plays a large role in transport of heat and water around the globe.

The land surface component, which simulates surface characteristics such as vegetation, snow cover, soil water, rivers, and carbon storing.

The ocean component, which simulates current movement and mixing, and biogeochemistry, since the ocean is the dominant reservoir of heat and carbon in the climate system.

The sea ice component, which modulates solar radiation absorption and air-sea heat and water exchanges.

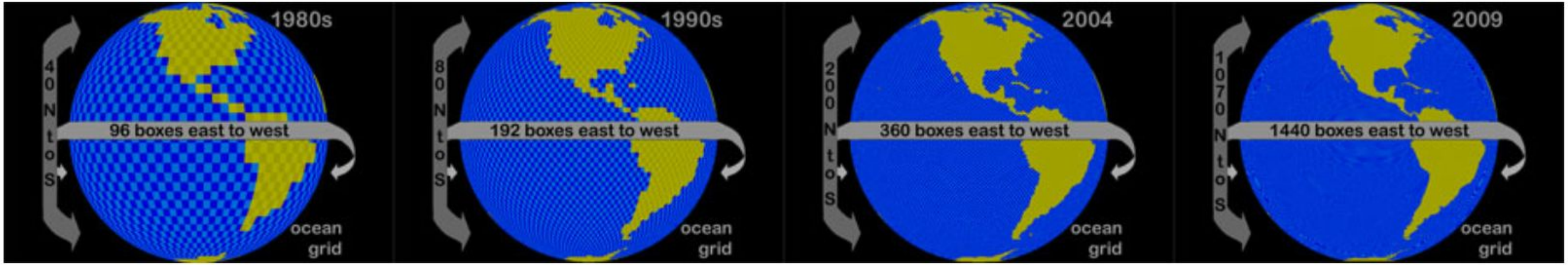


Washington, et al., 2008

gfdl.noaa.gov

Climate models

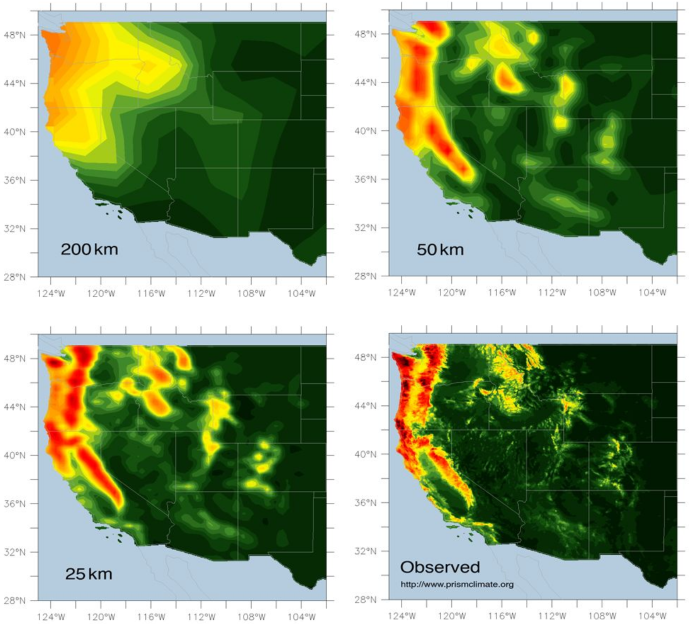
Climate models divide the globe into a three-dimensional grid of cells representing specific geographic locations and elevations. Each of the components (atmosphere, land surface, ocean, and sea ice) has equations calculated on the global grid for a set of climate variables such as temperature.



The spatial resolution of the grid depends on the the amount of computing power available.

gfdl.noaa.gov

Climate models



Current model resolution (200 km) compared to
high-resolution models (50 km and 25 km) and observed data

Observed data provided by PRISM Climate Group, Oregon State University.

Better spatial resolution leads to more accurate models.

gfdl.noaa.gov

Climate models are compute intensive!

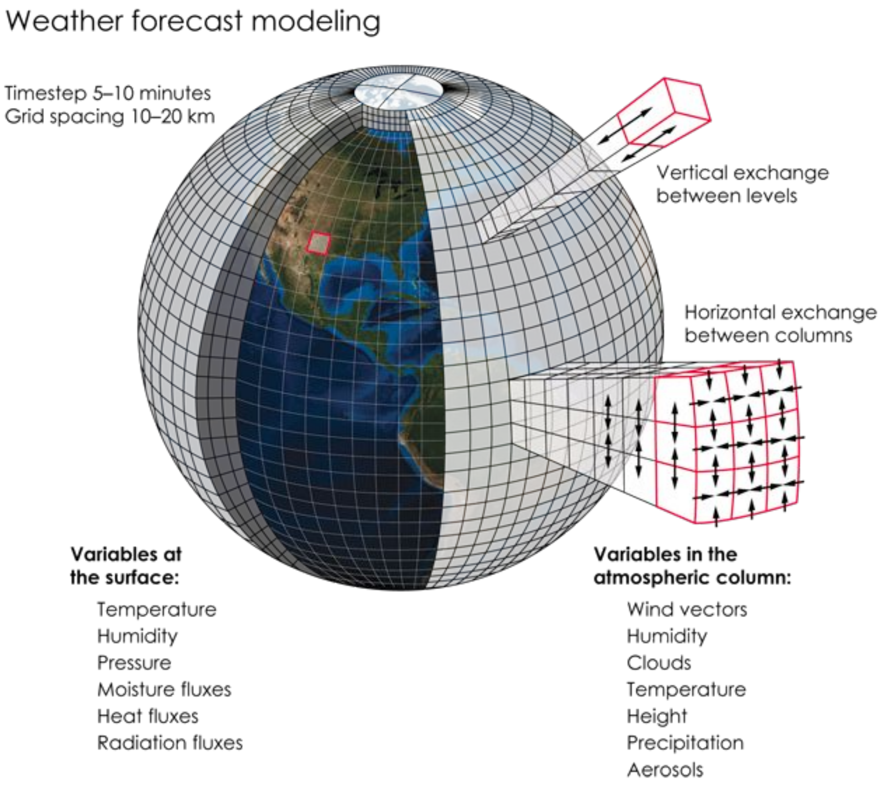
“A global climate model typically contains enough computer code to fill 18,000 pages of printed text; it will have taken hundreds of scientists many years to build and improve; and it can require a supercomputer the size of a tennis court to run.”

2-3 years of simulation time can take 1 day to run.

https://www.carbonbrief.org/qa-how-do-climate-models-work/

The Met Office Hadley Centre’s three new Cray XC40 supercomputers, for example, are together capable of 14,000 trillion calculations a second. The timelapse video below shows the third of these supercomputers being installed in 2017.

Physical simulations also power weather predictions



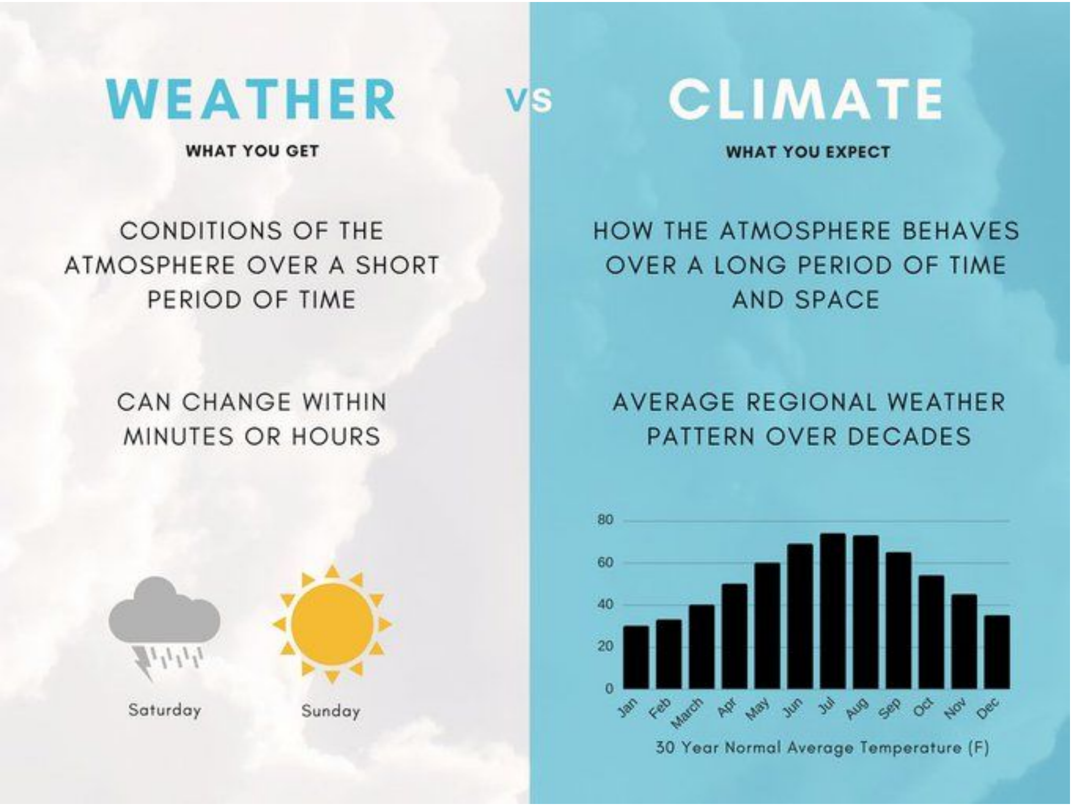
K. Cantner, AGI

Also use a 3-D grid

Smaller timestep: ~5 minutes for weather, closer to 30 min to an hour for climate.

Much shorter forecasting window: weeks instead of a century

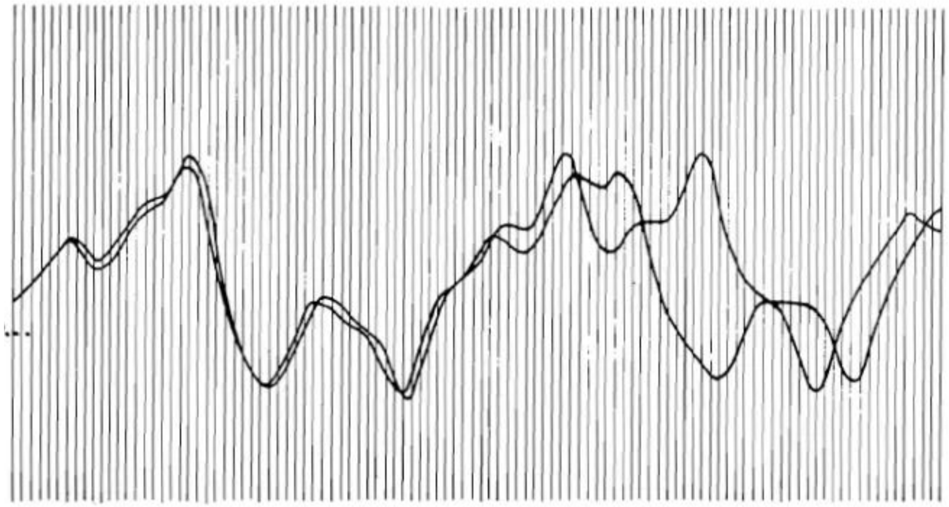
Climate and weather are different



Climate prediction can be successful over long time scales even as weather prediction on short time scales is difficult.

This is because climate variables are averages over space and time.

Limits to climate prediction



HOW TWO WEATHER PATTERNS DIVERGE. From nearly the same starting point, Edward Lorenz saw his computer weather produce patterns that grew farther and farther apart until all resemblance disappeared. (From Lorenz's 1961 printouts.)

The climate system is a "chaotic system" meaning that small initial differences can lead to large changes later.

Need to be able to run many simulations with different initial parameters to see the distribution of possible outcomes.

How can machine learning help?

Climate and Weather Model Emulators: Replace the physical simulation with a learned machine learning model that is faster to run (your homework)

Hybrid Models: Help run higher resolution models by learning to upsample low resolution values (the paper)

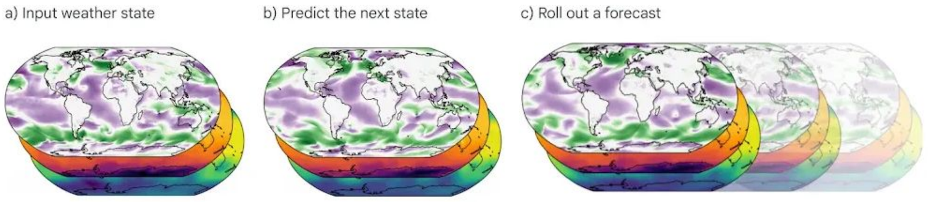
Weather model emulators are performing well as of late

GraphCast: AI model for faster and more accurate global weather forecasting

14 NOVEMBER 2023

Remi Lam on behalf of the GraphCast team

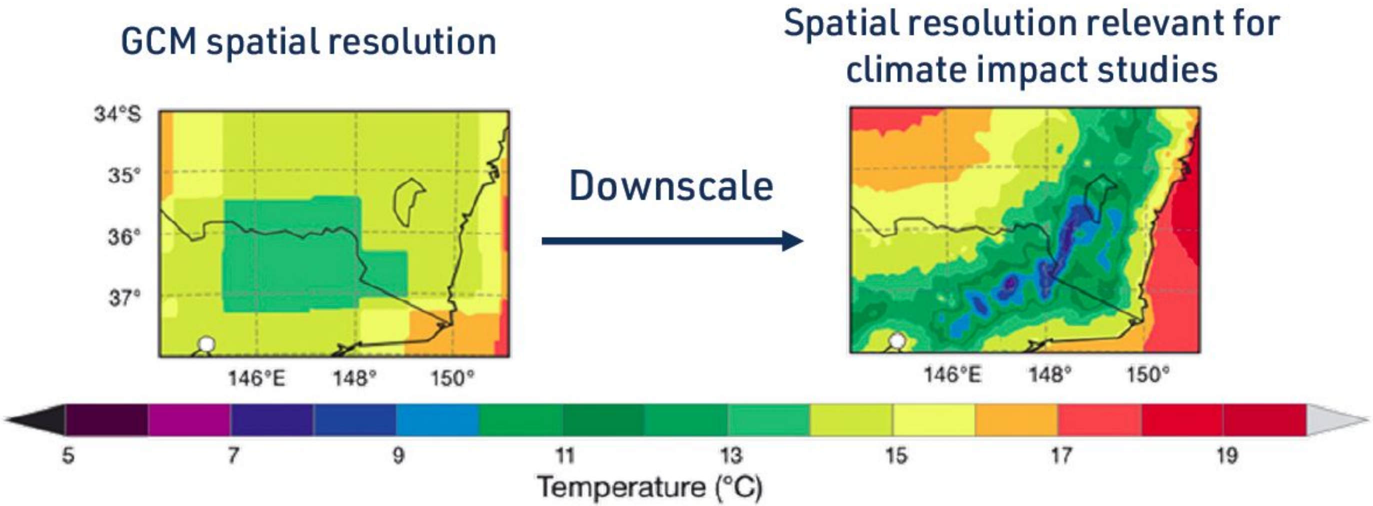
Our state-of-the-art model delivers 10-day weather predictions at unprecedented accuracy in under one minute



For inputs, GraphCast requires just two sets of data: the state of the weather 6 hours ago, and the current state of the weather. The model then predicts the weather 6 hours in the future. This process can then be rolled forward in 6-hour increments to provide state-of-the-art forecasts up to 10 days in advance.

GraphCast is trained on decades of historical weather data to learn a model of the cause and effect relationships that govern how Earth's weather evolves, from the present into the future.

Hybrid climate models

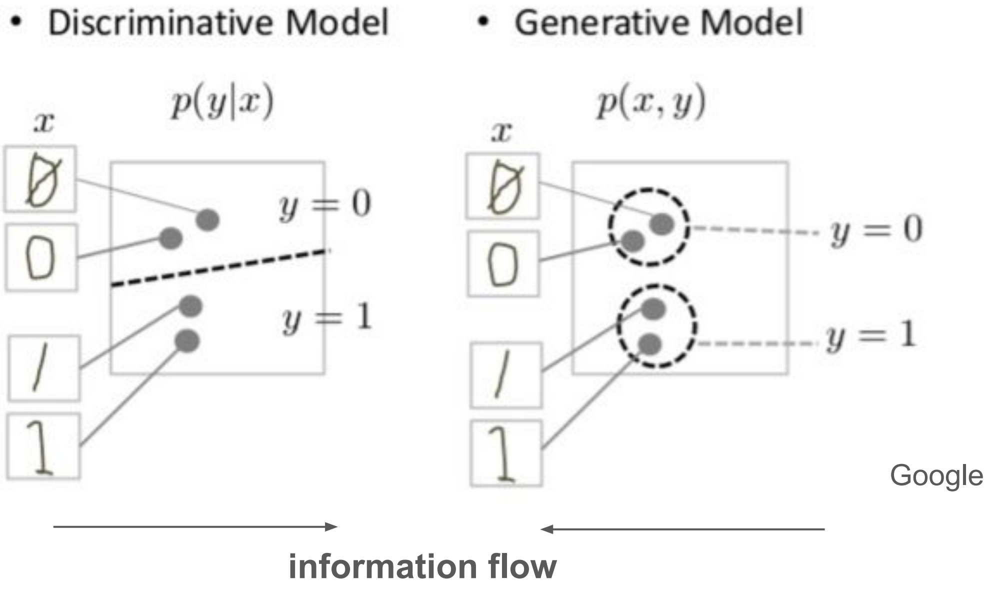


Statistical Downscaling: Learn a model that can project poor resolution information into a higher resolution ("downscaling" = going to a lower spatial scale, "statistical" = learned from data). Also known as "super-resolution".

This is a "generative" modeling problem

NARCLiM

Generative models



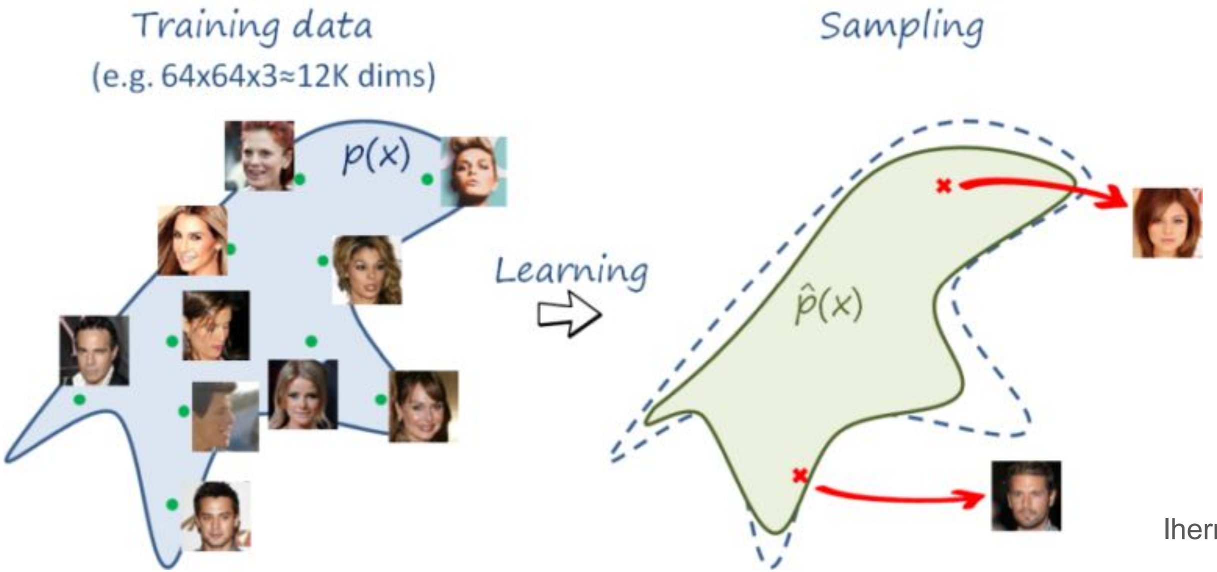
Discriminative: model learns to map a high-dimensional input to a lower dimensional label

Generative: model learns to create a high-dimensional output that is a sample pulled from the distribution learned from training data

thispersondoesnotexist.com

Images of faces created by a generative model trained on pictures of human faces

Generative models

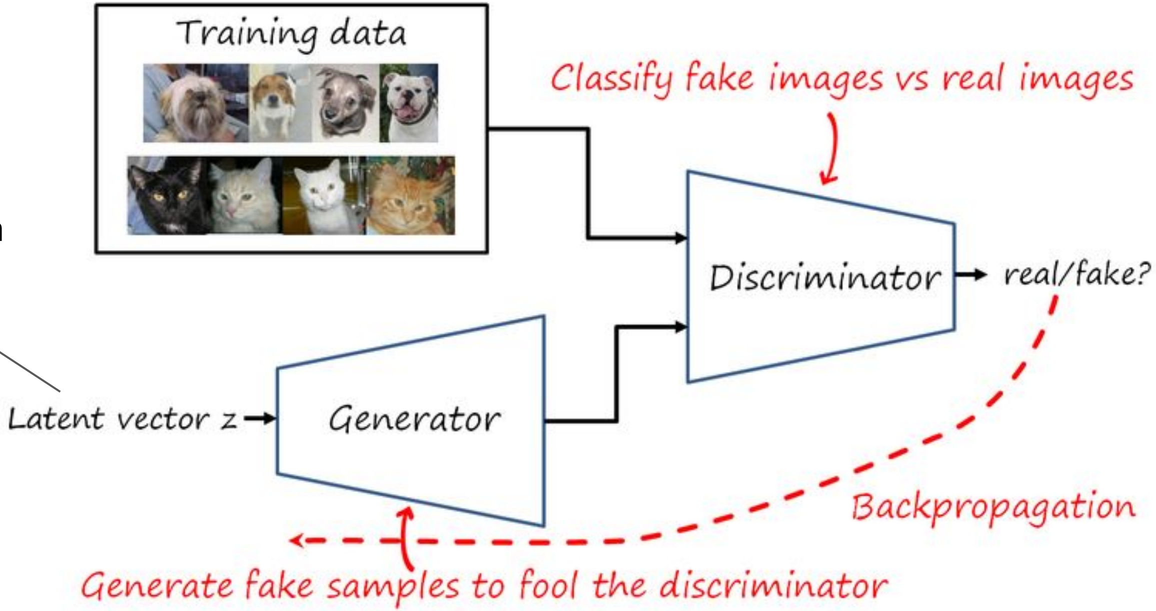


lherranz.org

The model learns what patterns of pixel values are statistically likely in face images. It can then generate a new image that has a high probability of being a face.

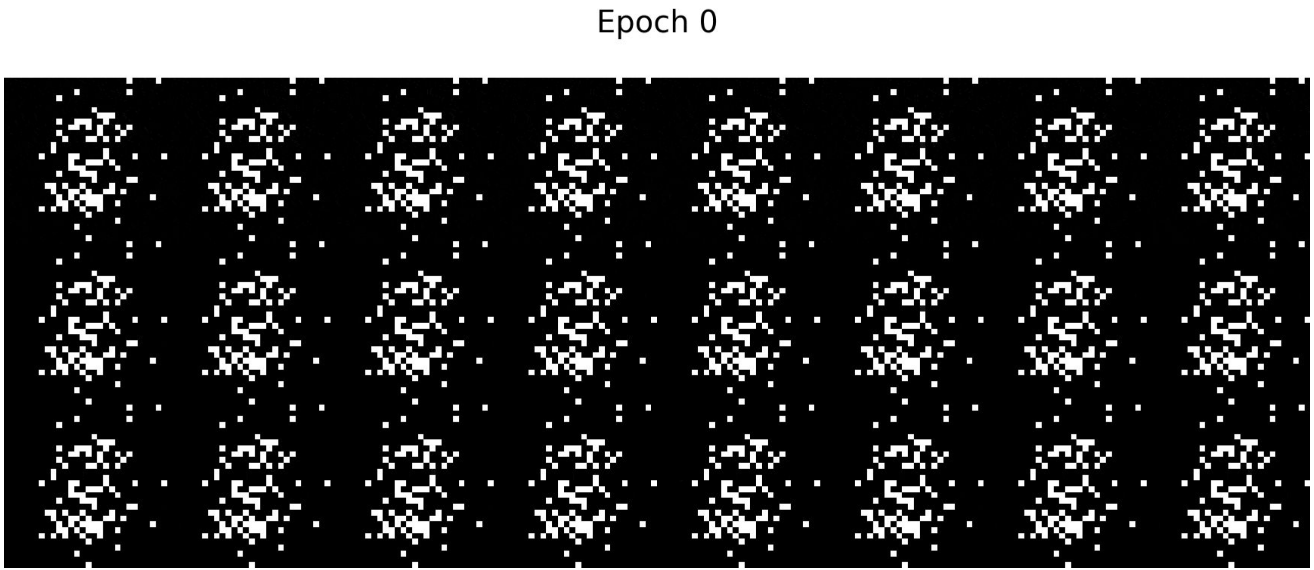
Generative adversarial networks

Usually random values, but can also include, e.g., category label



The specific model used is a GAN: Generative Adversarial Network. By jointly training the generator and discriminator (with different loss functions), the generator learns to create convincing fake images

Generative adversarial networks (GANs)

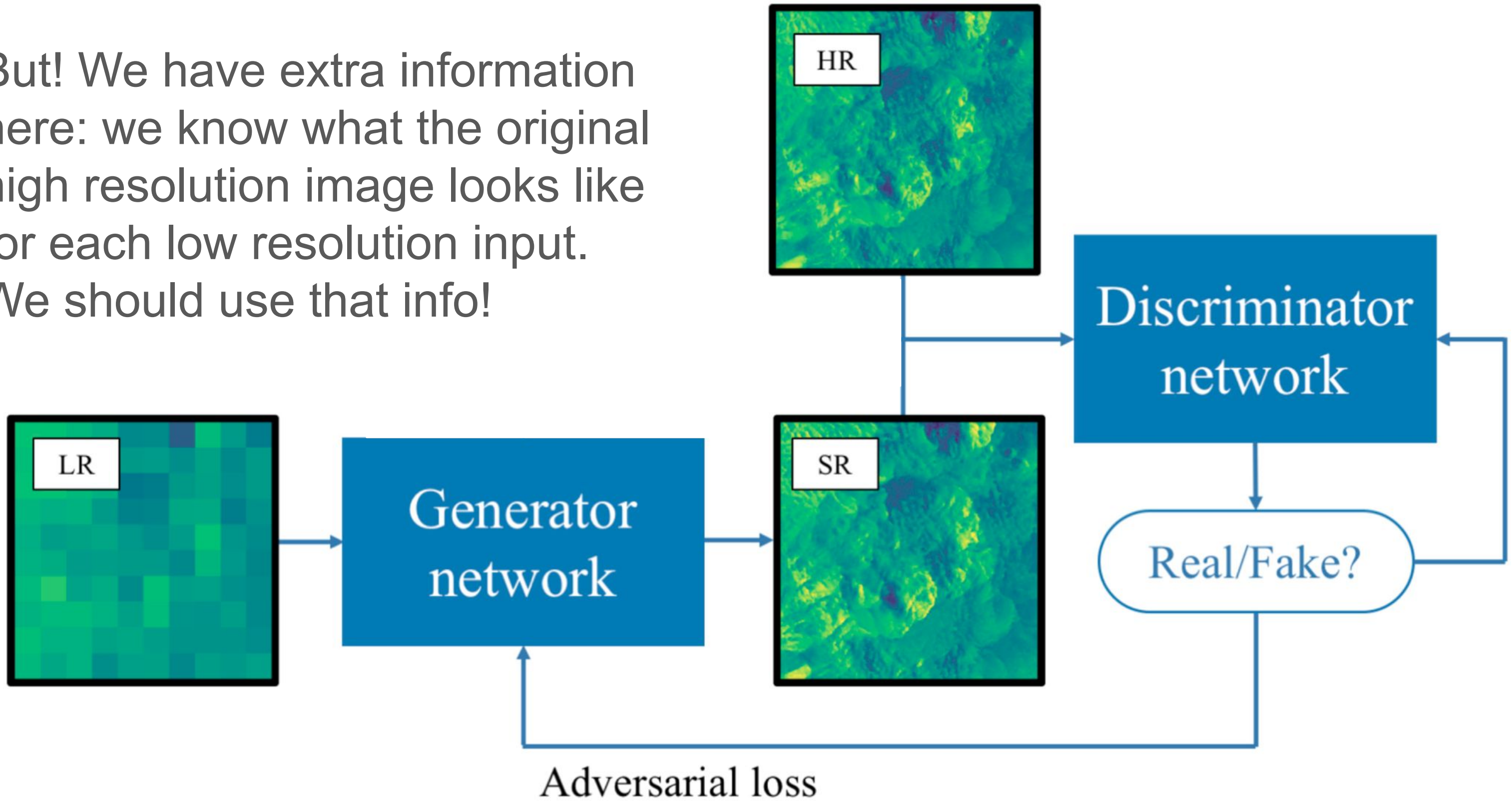


Here, the generator learned the statistics of hand-written digits.

https://sthalles.github.io/intro-to-gans/

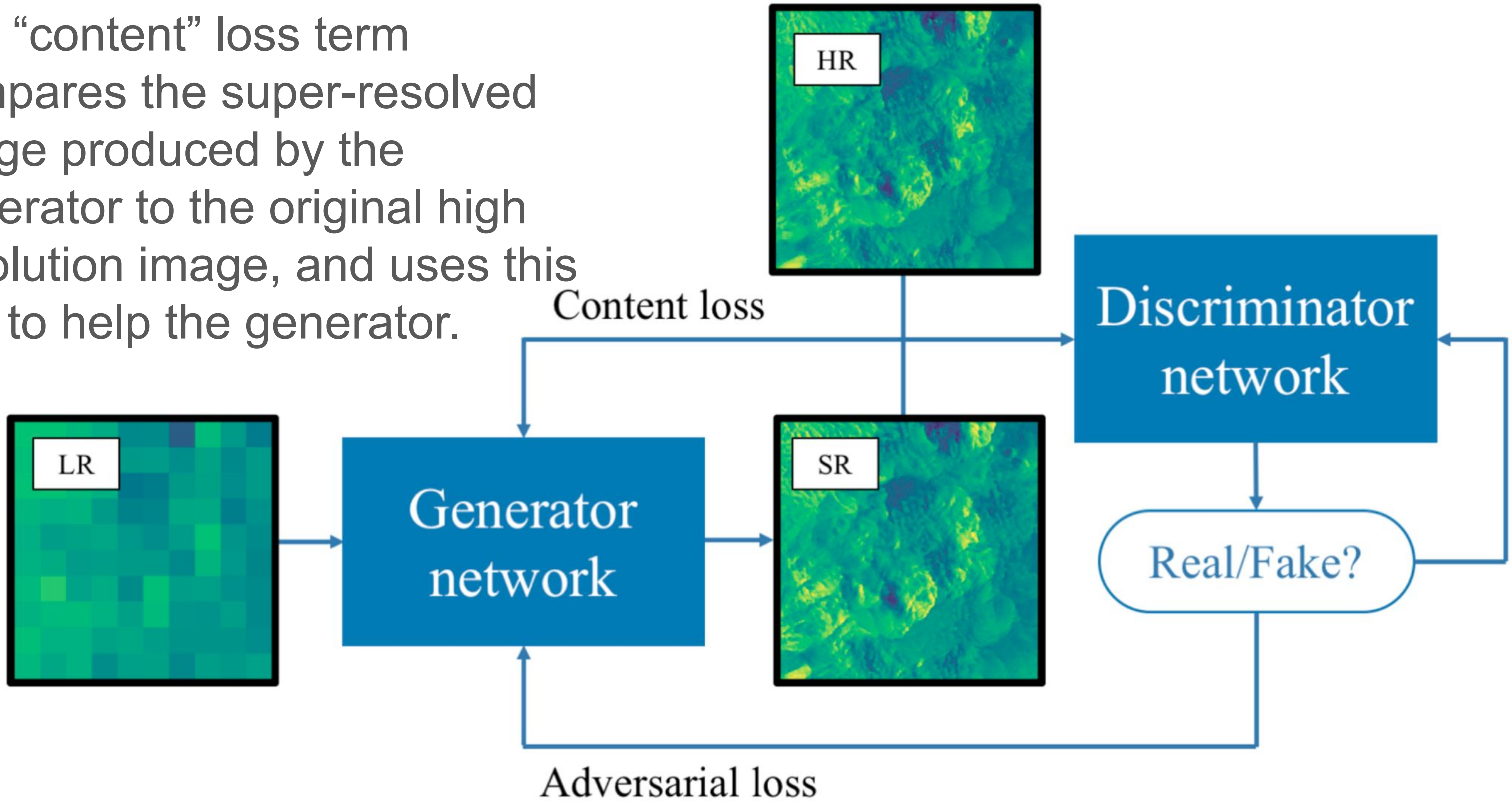
Generative models can be used to solve super resolution

But! We have extra information here: we know what the original high resolution image looks like for each low resolution input. We should use that info!



Generative models can be used to solve super resolution

The “content” loss term compares the super-resolved image produced by the generator to the original high resolution image, and uses this info to help the generator.



Loss function for a super-resolution generator

$$l^{SR} = \underbrace{l_{\mathrm{X}}^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

Want to:

Recreate the true specific high resolution image associated with given low resolution image

Trick the discriminator into thinking this is a real high resolution image

Loss function for a super-resolution generator

$$l^{SR} = \underbrace{l_{\mathrm{X}}^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

Want to:

Recreate the true specific high resolution image associated with given low resolution image

Trick the discriminator into thinking this is a real high resolution image

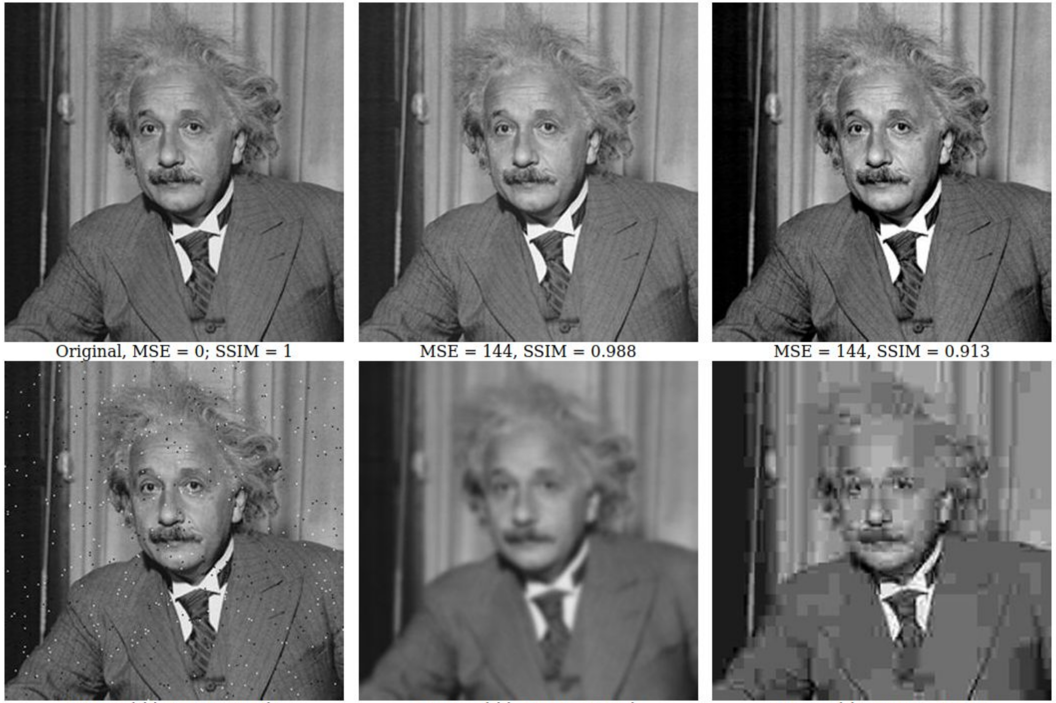
What mathematical function should this be?

How do we compare two images?

Pixelwise
mean-squared error

$$MSE = \frac{1}{n}\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y(i,j) - \hat{Y}(i,j))^2$$

where m,n are spatial dimensions



Original, MSE = 0; SSIM = 1

MSE = 144, SSIM = 0.988

MSE = 144, SSIM = 0.913

MSE = 144, SSIM = 0.840

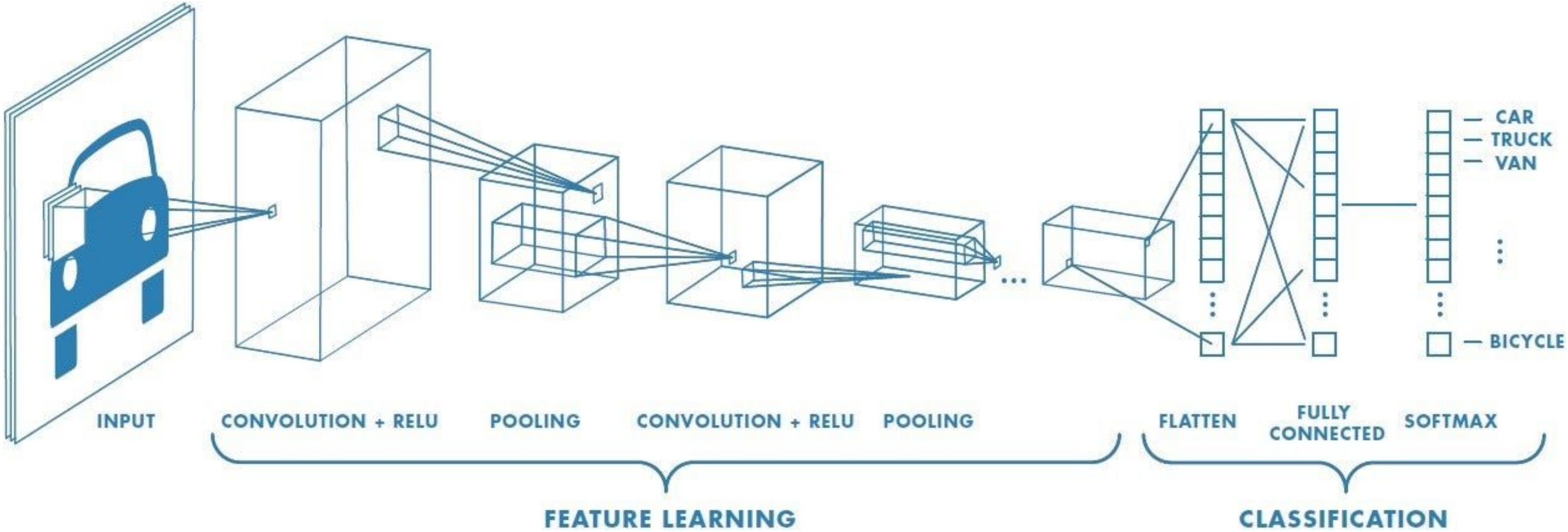
MSE = 144, SSIM = 0.694

MSE = 142, SSIM = 0.662

MSE doesn't necessarily capture what we care about!

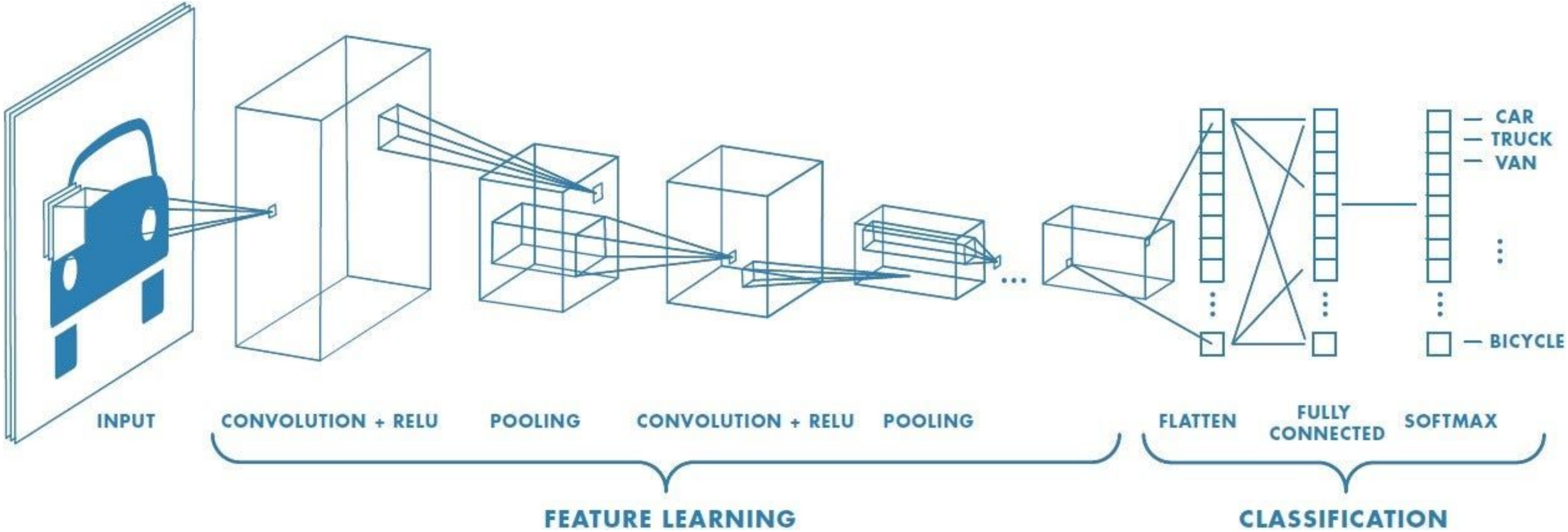
cns.nyu.edu/~lcv/ssim/

Can CNNs help us compare images?



Convolutional neural networks can learn to represent images in a way that aligns more with our perceptual experience

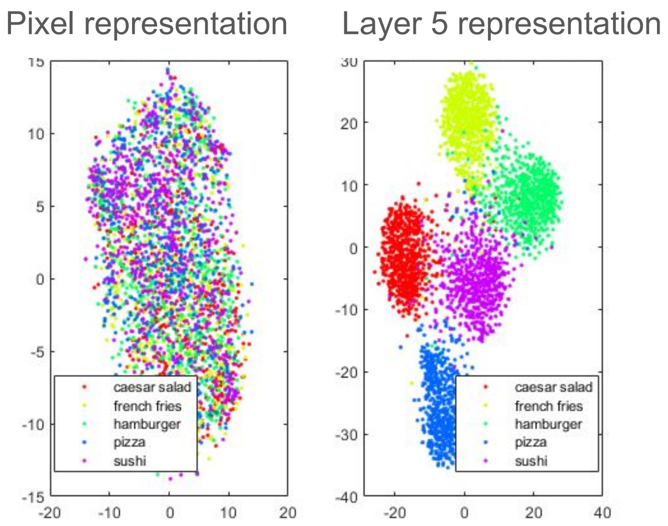
Can CNNs help us compare images?



Representations = transformations of the original data. In ANNs, the activity of units at different layers in the network can be used as different ways of representing the image.

Can CNNs help us compare images?

Dimensionality Reduction applied to different layers of a CNN trained on food images:

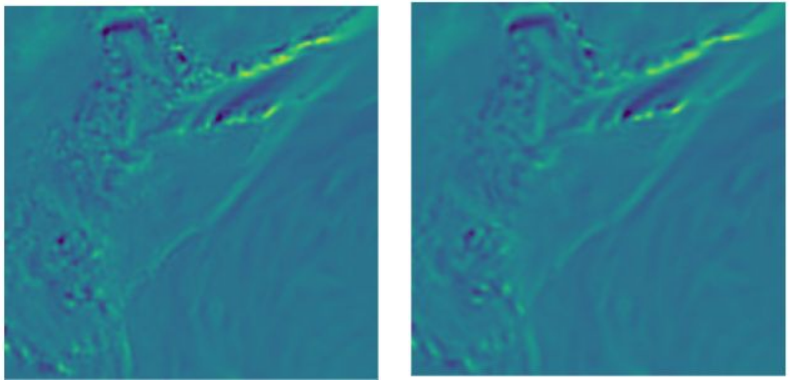


Representations = transformations of the original data. In ANNs, the activity of units at different layers in the network can be used as different ways of representing the image.

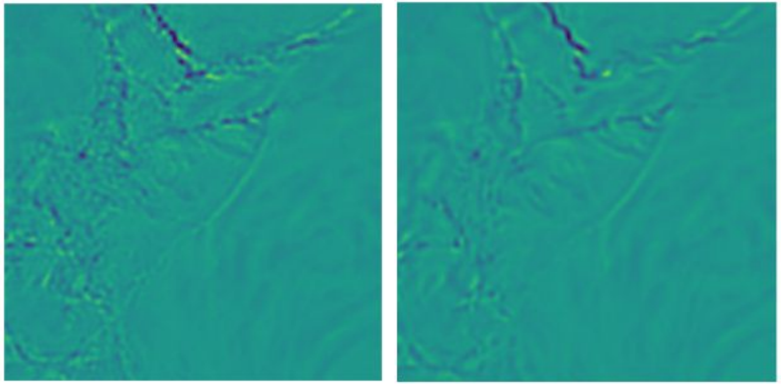
mathworks

But a CNN trained to classify food won't capture what we care about for super-resolving climate models...

Paper is focused on downscaling atmospheric variables:



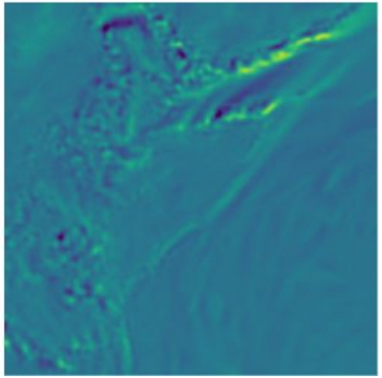
How similar are there two "vorticity spatial fields" (spinning atmospheric wind patterns)?

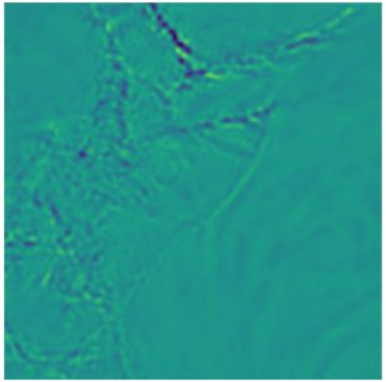


How similar are there two "divergence spatial fields" (downward atmospheric wind patterns)?

Need to train a CNN on atmospheric images

But what should the loss function be?





Label: ?? ??

Self-supervised learning doesn't require traditional labels

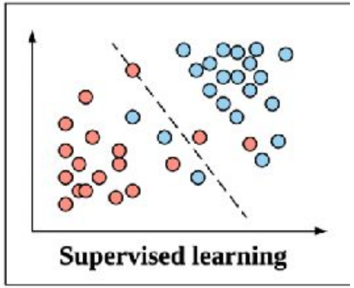
Many different ways to learn representations:

All data points need separate labels

A few data points are labeled

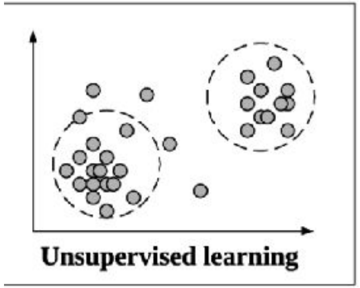
Other aspects of the data are used to create "labels"

No data points are labeled

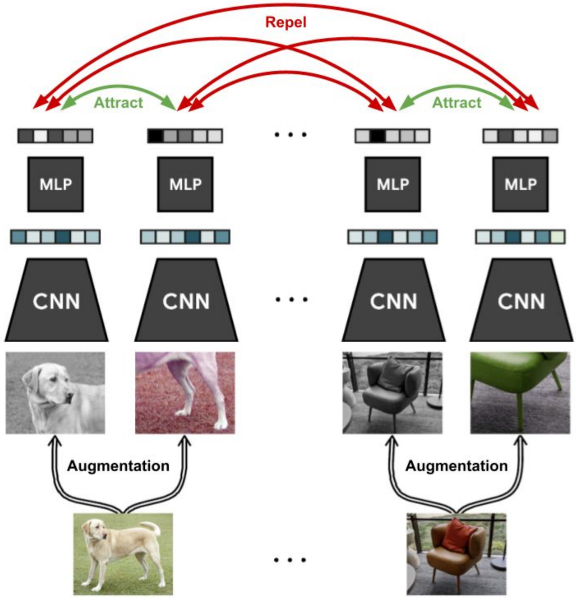


Semi-supervised learning

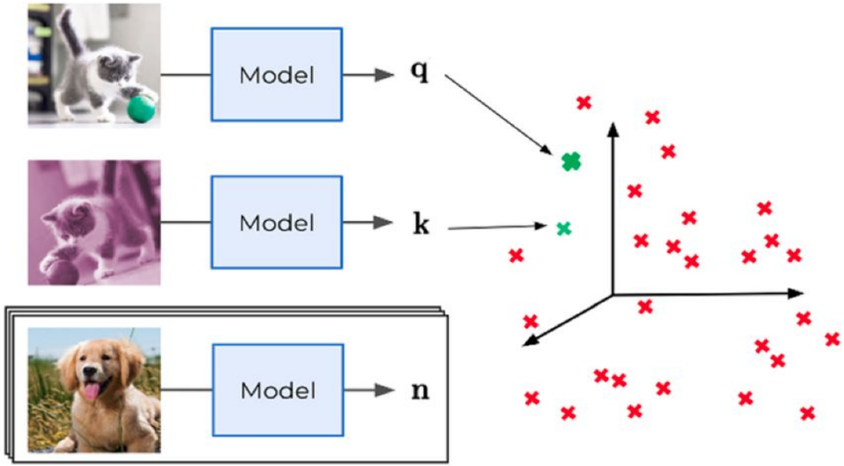
Self-supervised learning



Contrastive learning: type of self-supervised learning



Ting Chen et al



Kumar et al

The model should learn to represent similar things similarly and different things differently. One possible way to do this is add augmentations to an image and tell the network to label the different versions of the original as the same, but versions of other images differently.

Overview of the paper’s method:

- Train a self-supervised model on images of atmospheric data
- Use the representation from the self-supervised model as a measure of image similarity
- Include this measure of image similarity in the loss function of a GAN, which is trained to downscale atmospheric data images.
- Evaluate how well the GAN performs using other metrics relevant to atmospheric data