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 27 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 so it is unclear how useful it is in the real world

Climate Change in the News

White House Failed to Comply With Court Order, Judge Rules

The federal judge in Rhode Island said the Trump administration had failed to comply with his order unfreezing billions of dollars in federal grants.

Share full article

Judge McConnell had previously ordered the White House to unfreeze federal funds locked up by the White House budget office. A memo from that office had demanded that billions of dollars in grants be held back until they were determined to be in compliance with President Trump's priorities and ideological agenda.

On Friday, 22 Democratic attorneys general went to Judge McConnell to accuse the White House of failing to comply with his earlier order. The Justice Department responded in a filing on Sunday that money for clean energy projects and transportation infrastructure, allocated to states by the Inflation Reduction Act and the bipartisan infrastructure bill, was exempt from the initial order because it had been paused under a different memo.

Judge McConnell's ruling on Monday explicitly rejected that argument.

Automotive News

Inflation Reduction Act changes would threaten EV investments, jobs in pro-Trump states

Hannah Lutz
Wed, February 12, 2025 at 7:00 AM EST · 7 min read

As President Donald Trump considers changes to the Inflation Reduction Act, more than $100 billion in electric vehicle manufacturing investment and about 84,000 jobs are at stake in states he won in November's election.

Ninety percent — or $105 billion — of the EV manufacturing investment announced since the passage of the act went to Republican-leaning states, according to an Automotive News analysis of Atlas Public Policy data.

The most investment is slated for Georgia, North Carolina, South Carolina, Michigan and Indiana — with more than $12.5 billion each. By comparison, just $11 billion of the announced investments went to states that voted for Democrat Kamala Harris, Trump's opponent in the November election.

Some Republican members of Congress are pushing to keep the incentives to bolster the economies of their districts. Many repeated a phrase that has circulated in Republican lawmakers' recent discussions of the act: Refine the law with a scalpel, not a sledgehammer.

Republican Rep. Buddy Carter of Georgia said the act spurs economic growth that will make the U.S. competitive and secure.

Climate Change in the News

CLIMATE

India wants to embrace nuclear power. To do it, it'll need a lot of time and money

BY SIBI ARASU

Updated 10:10 PM EST, February 11, 2025

Share

BENGALURU, India (AP) — India wants more nuclear power, has pledged over $2 billion toward research and will change laws to boost investment to do it.

The pledges were made by India's finance minister earlier this month as part of a plan to expand electricity generation and reduce emissions. Nuclear power is a way to make electricity that doesn't emit planet-warming gases, although it does create radioactive waste. India is one of the world's biggest emitters of planet-heating gases and over 75% of its power is still generated by burning fossil fuels, mostly coal. India wants to install 100 gigawatts of nuclear power by 2047 — enough to power nearly 60 million Indian homes a year.

Energy experts say that for the world to move away from carbon-polluting fuels like coal, oil and gas, sources like nuclear that don't rely on the sun and the wind — which aren't always available — are needed. But some are skeptical about India's ambitions as the country's nuclear sector is still very small, and negative public perceptions about the industry remain.

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 be the attendance taker. Have them go to this Google Form and enter the netIDs of all members of the group who are present.

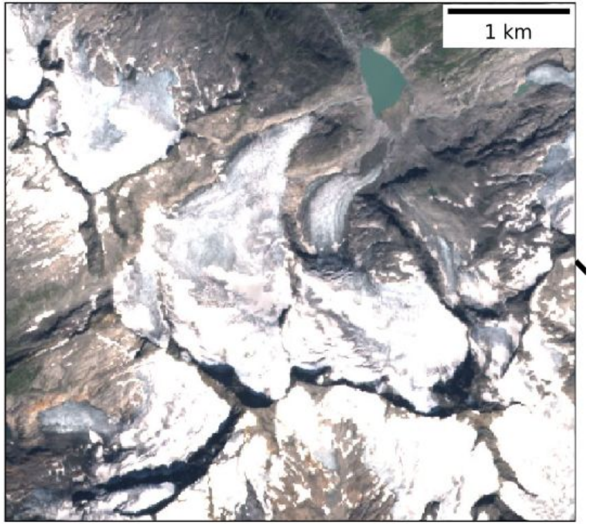
https://forms.gle/SsipLSQjwQCneQvV9 (link is also in Brightspace under Syllabus content)

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.

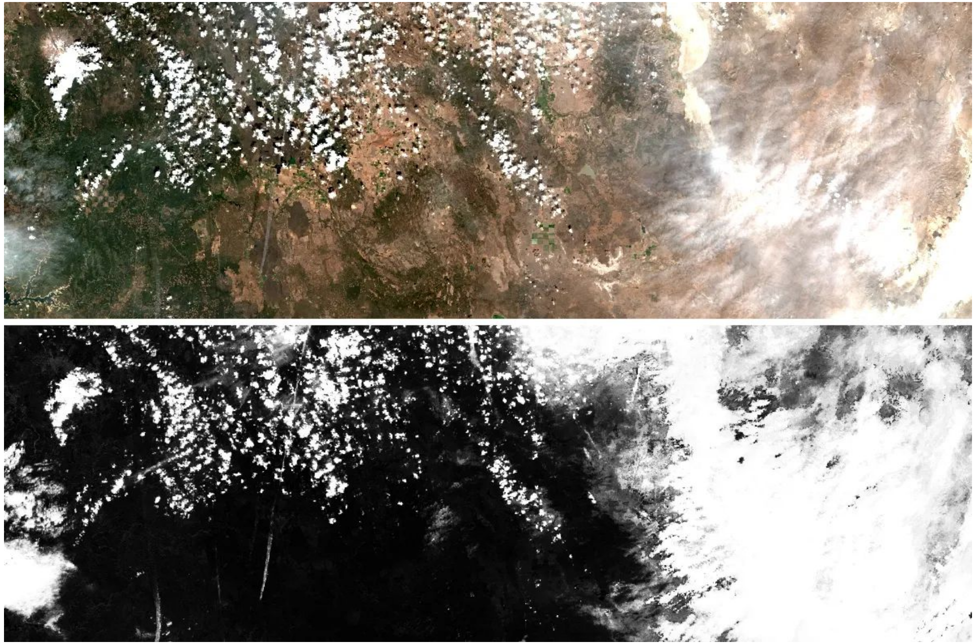
But, because there are many large glaciers, it does not significantly impact the total amount of glacier cover included in the data (~95%), which is arguably the more important measure here.

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

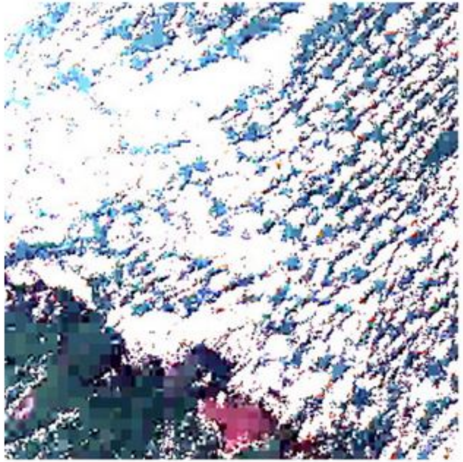
Sentinel-2 Cloud Coverage



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

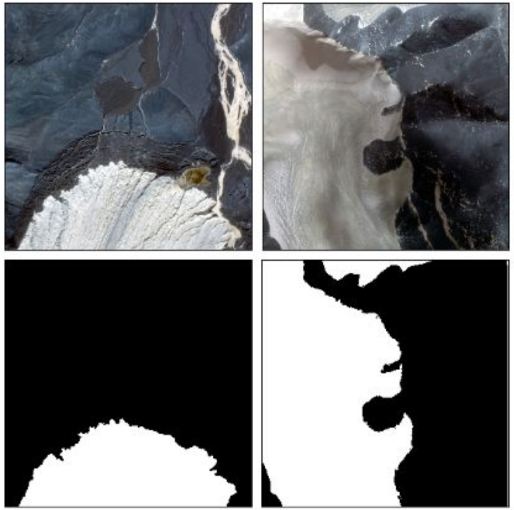
What loss function was used to train the model? Why is this appropriate for the problem?

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



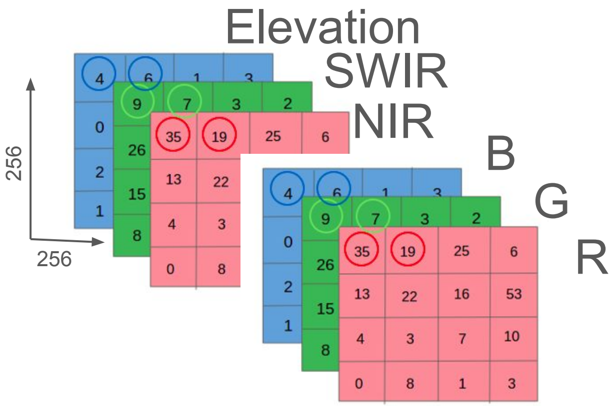
Chu et al

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

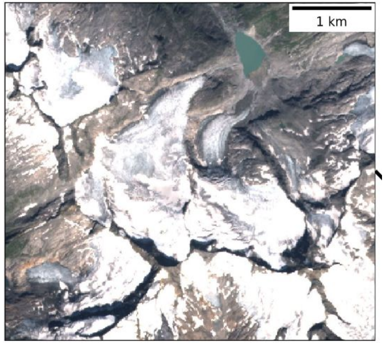
Larger images are cut into smaller ones

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.

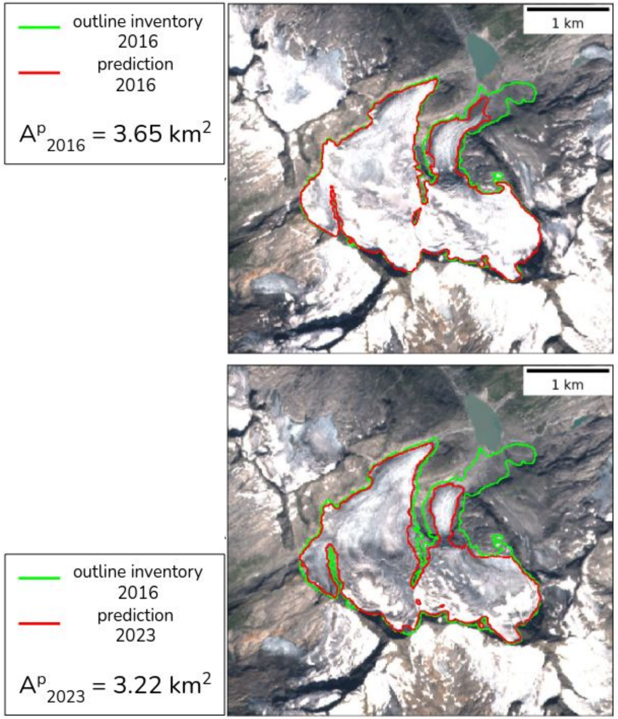
Discussion Question 6

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 2016 (green). 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

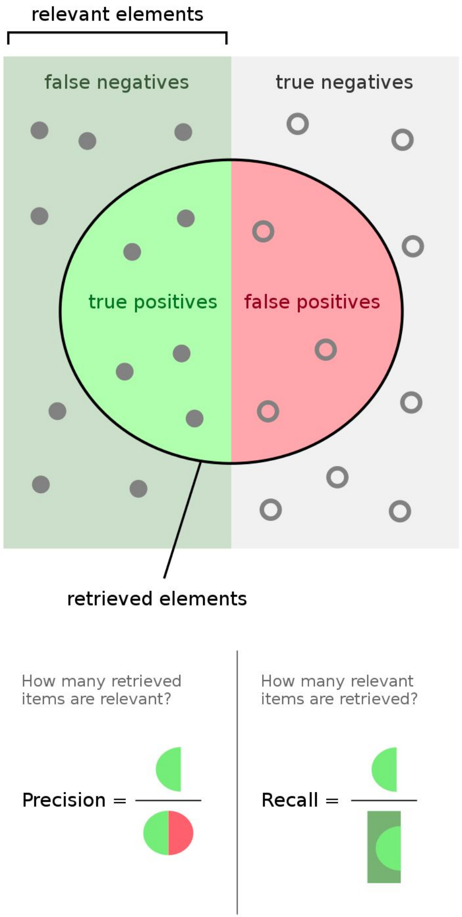
Which does the model do better at: labeling glacier pixels as glaciers or not labeling non-glacier pixels as glaciers?

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).



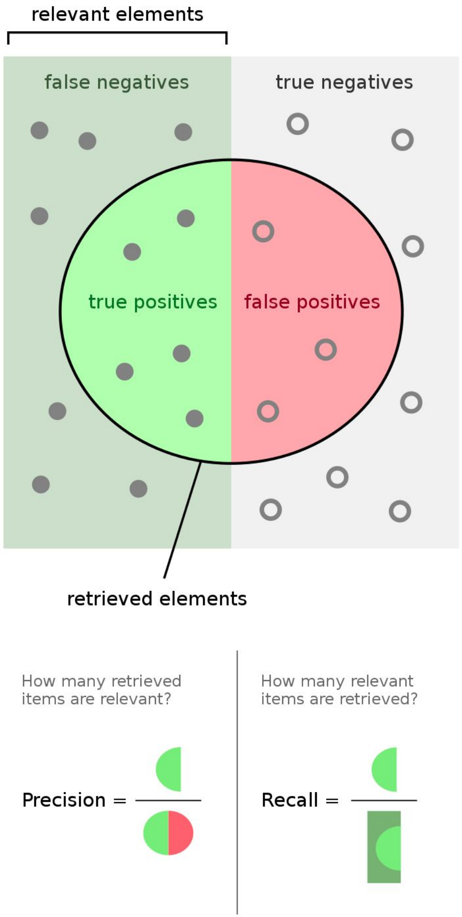
Precision and Recall

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	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	0.92 ± 0.03	0.89 ± 0.06	0.90 ± 0.03

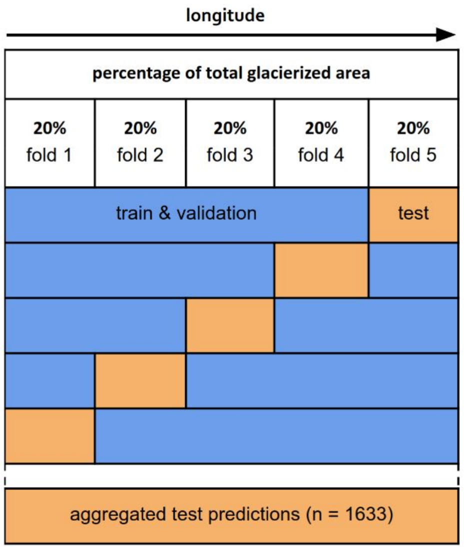


Figure A1: Cross-validation scheme with a geographic split

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, representation learning, 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

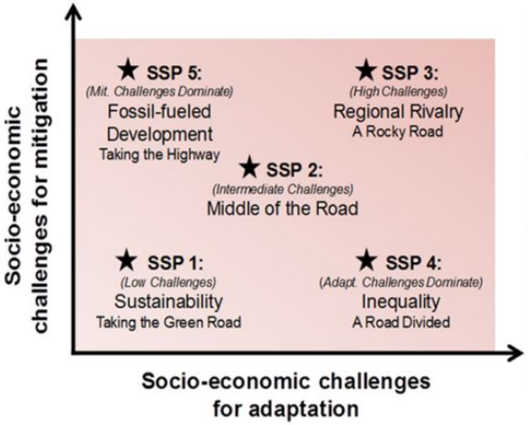
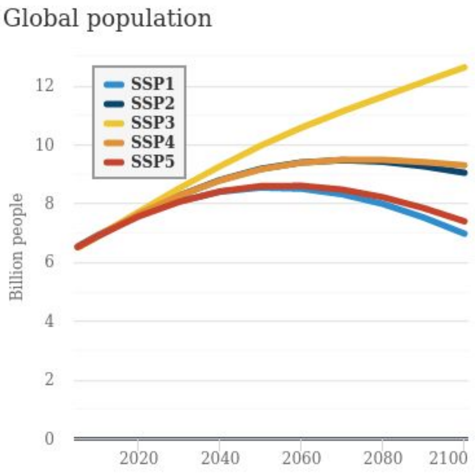
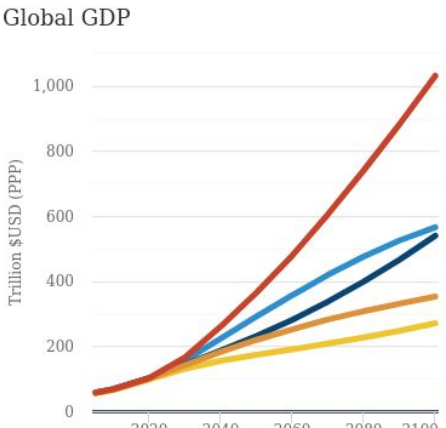


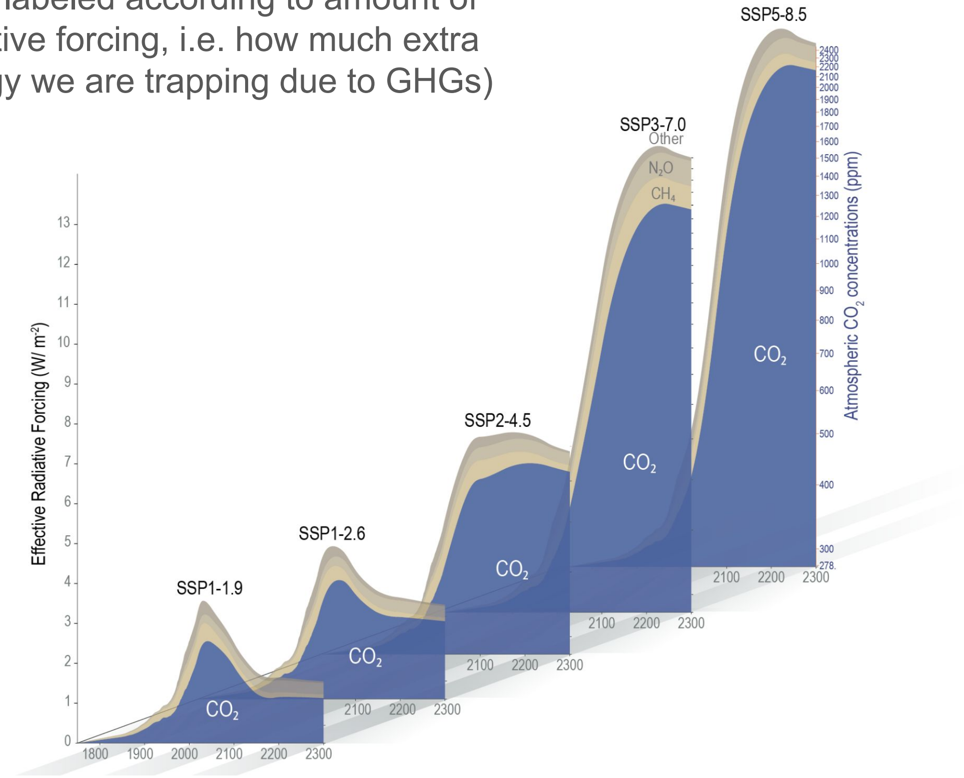
Fig. 1 Overview of SSPs
(Narratives in O'Neill et al., 2016, Glob Env Change, online first)
SSP1: low challenges for mitigation (resource efficiency) and adaptation (rapid development)
SSP3: high challenges for mitigation (regionalized energy / land policies) and adaptation (slow development)
SSP4: low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

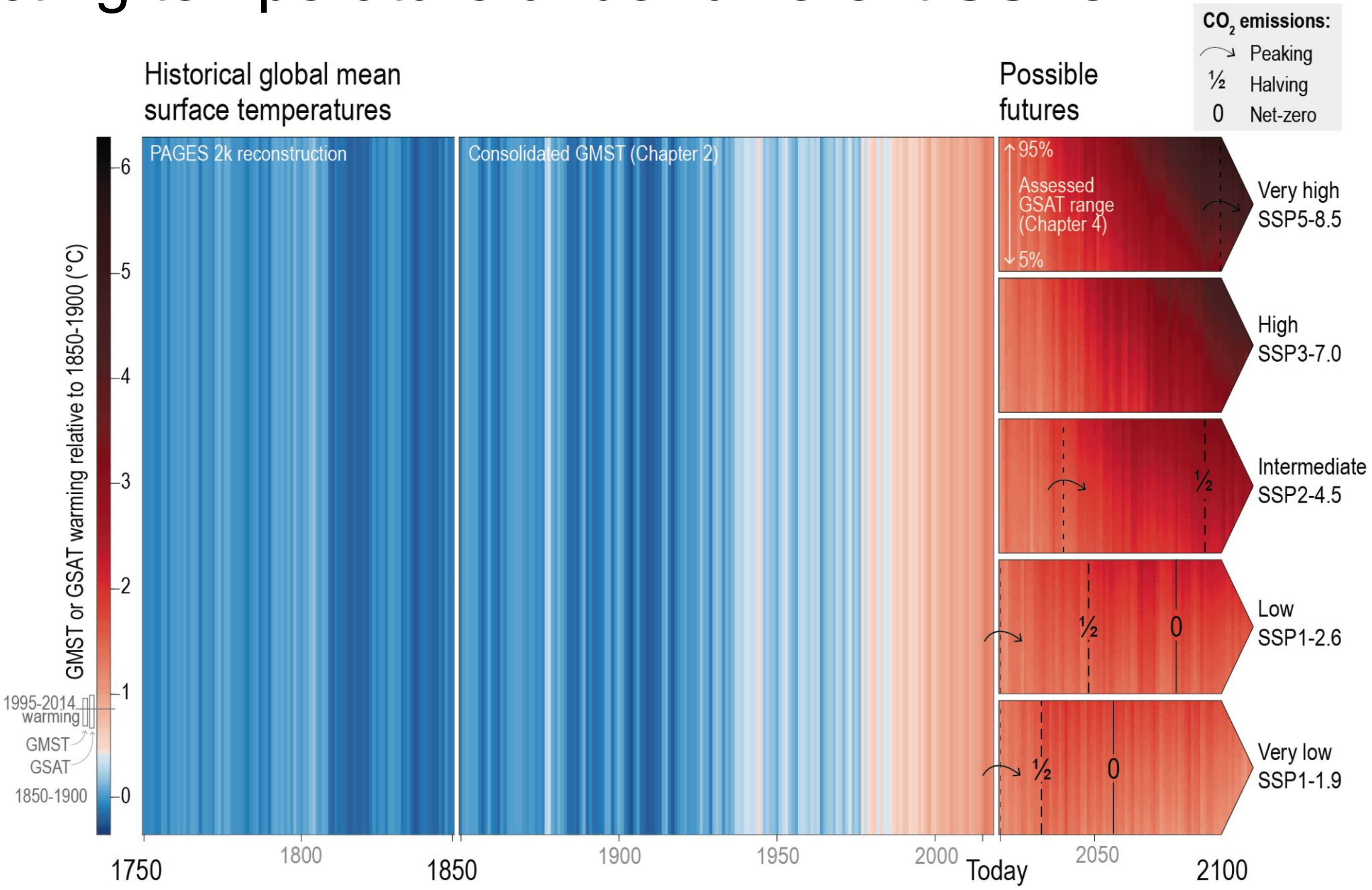




(Also labeled according to amount of radiative forcing, i.e. how much extra energy we are trapping due to GHGs)

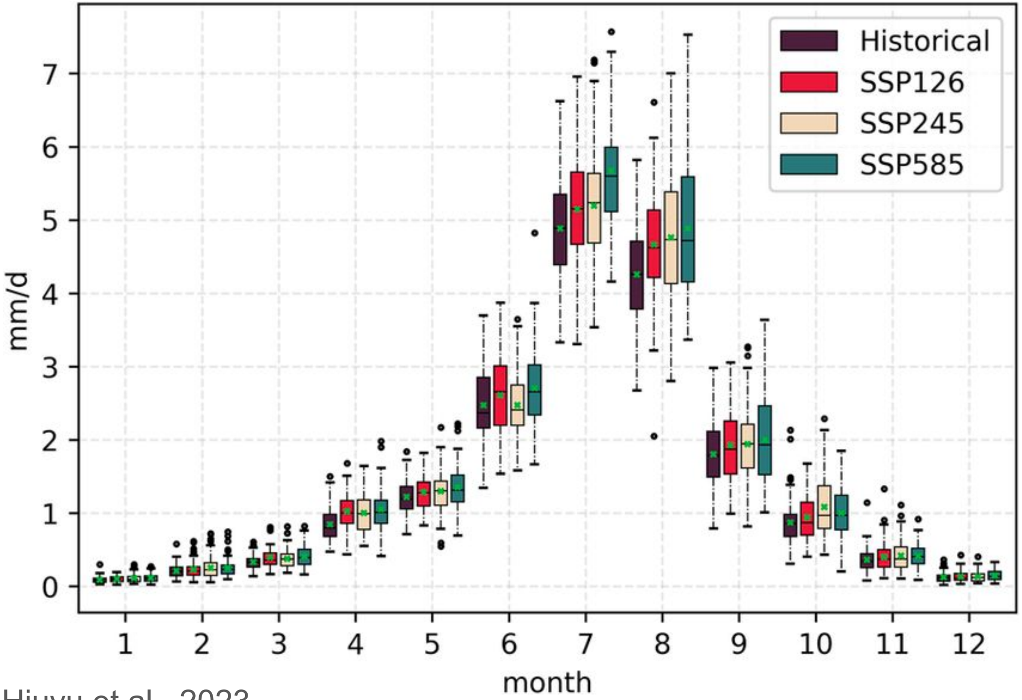


Predicting temperature under different SSPs



Predicting precipitation under different SSPs

Monthly precipitation



Hiuyu et al., 2023

How do we make these predictions?

Climate simulation models!

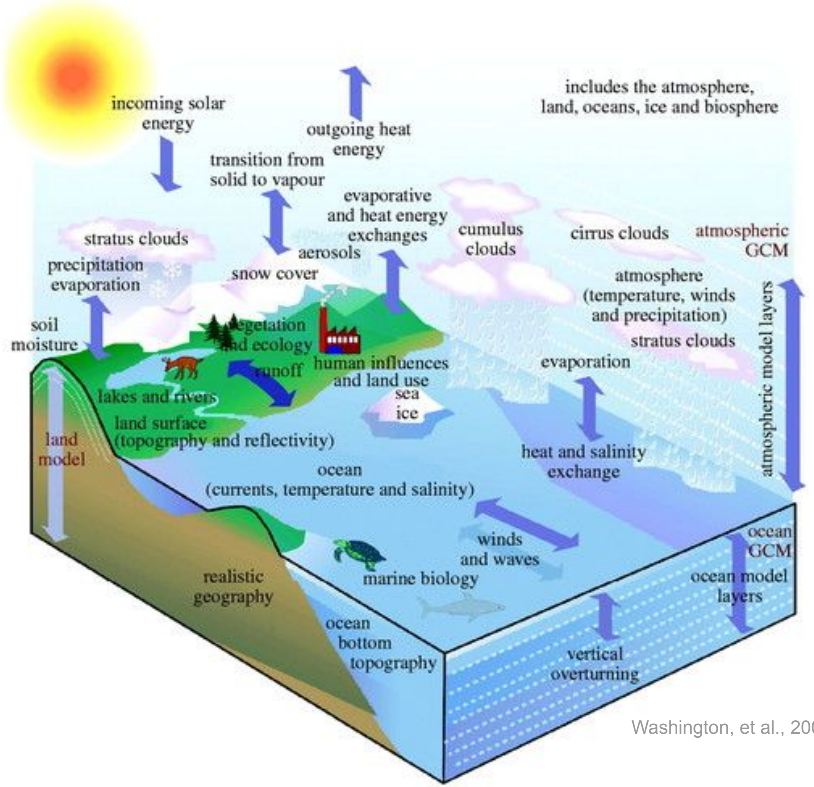
A **global climate model** (GCM) is a complex mathematical representation of the major climate components and their interactions. The main climate system components treated in a climate model are:

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.

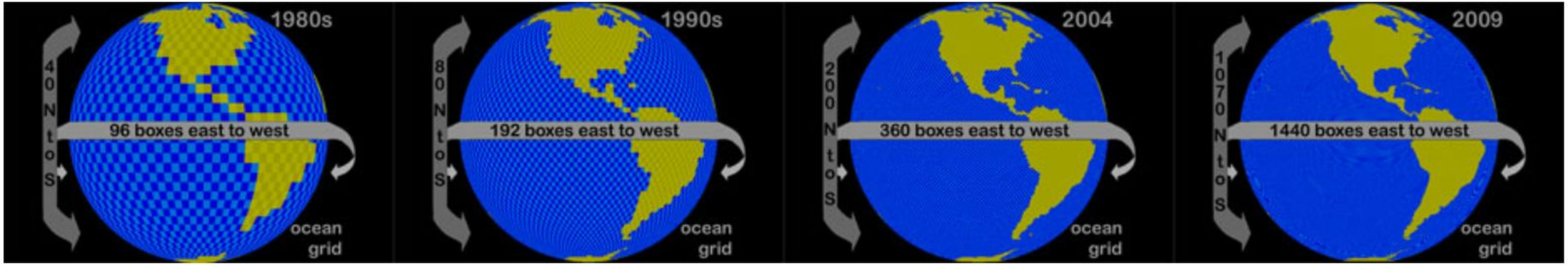


Washington, et al., 2008

gfdl.noaa.gov

Climate models

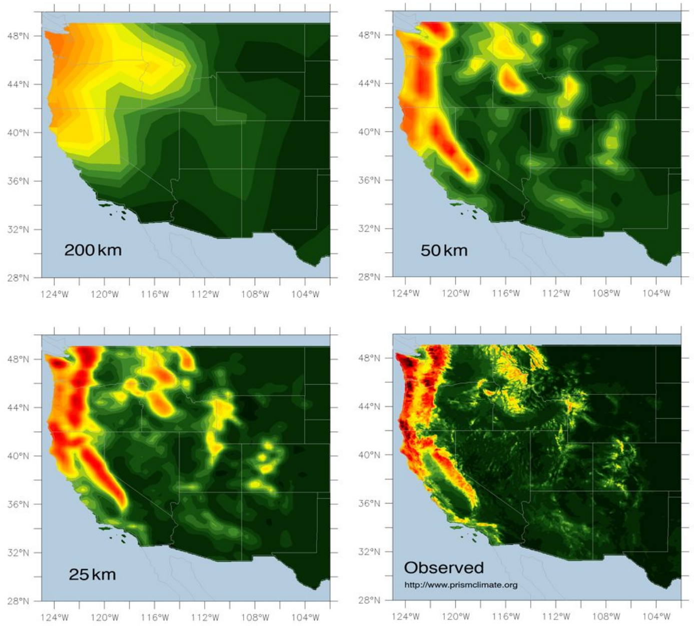
Climate models divide the globe into a three-dimensional grid of cells representing specific geographic locations and elevations. Each of the components (atmosphere, land surface, ocean, and sea ice) has equations calculated on the global grid for a set of climate variables such as temperature.



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!

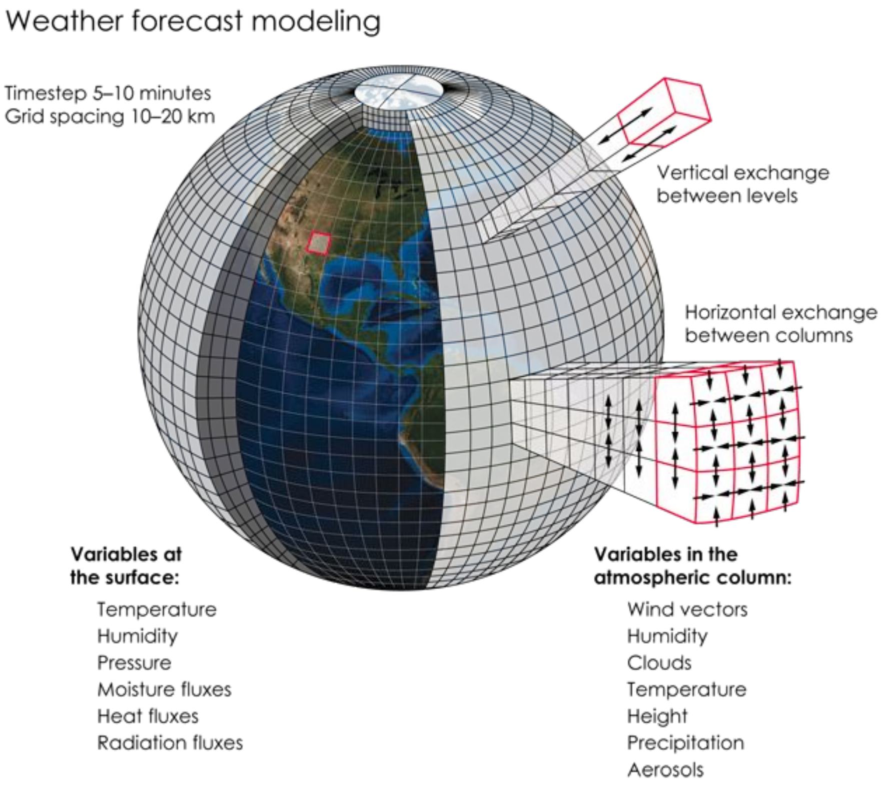
"A global climate model typically contains enough computer code to fill 18,000 pages of printed text; it will have taken hundreds of scientists many years to build and improve; and it can require a supercomputer the size of a tennis court to run."

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



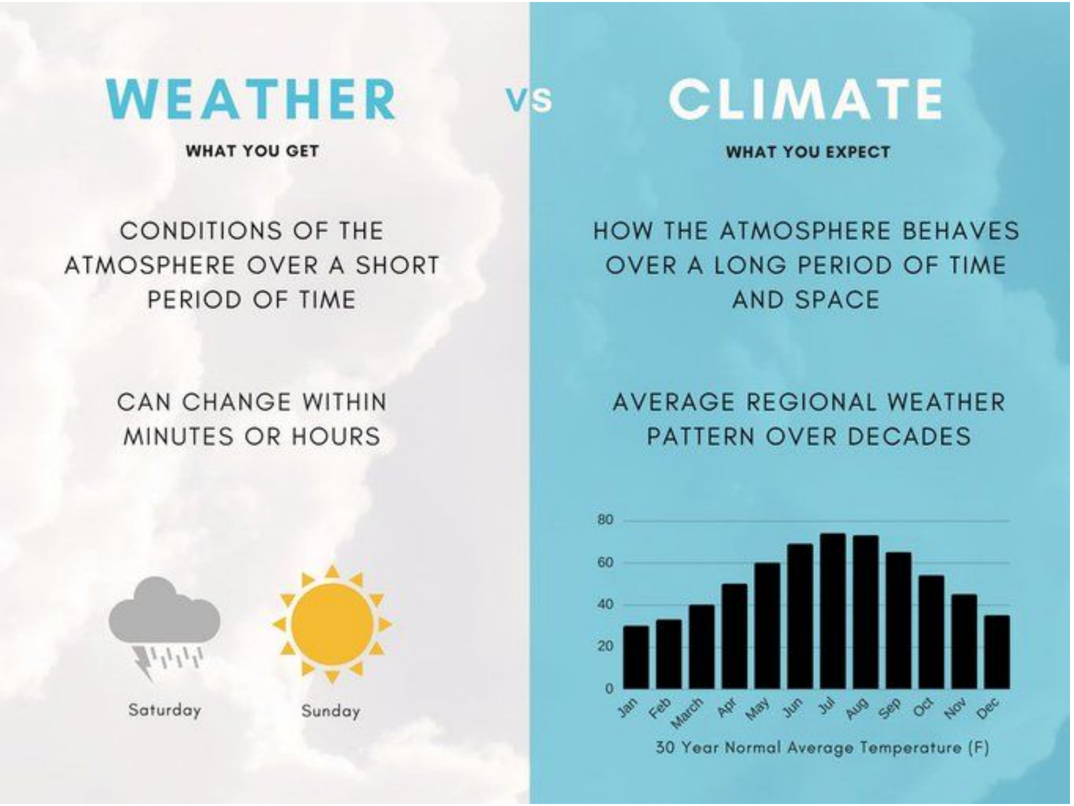
K. Cantner, AGI

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

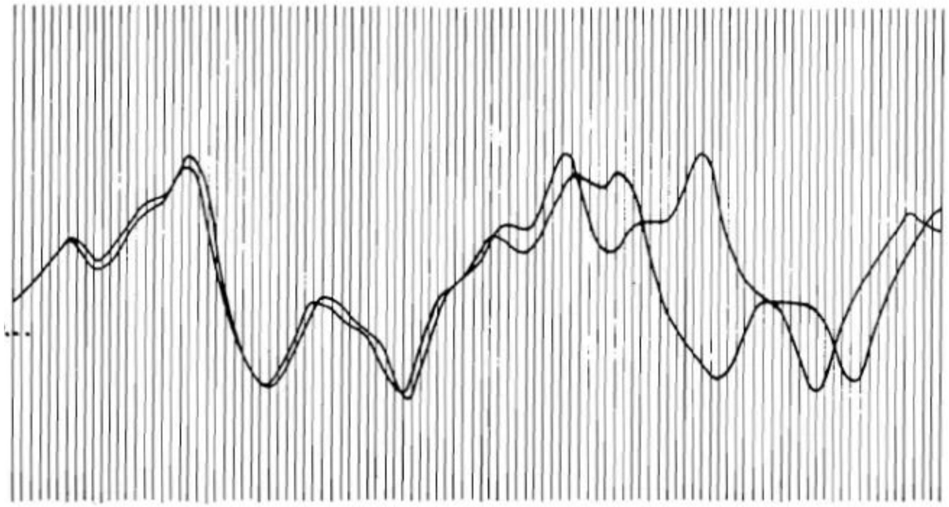
Climate and weather are different



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.

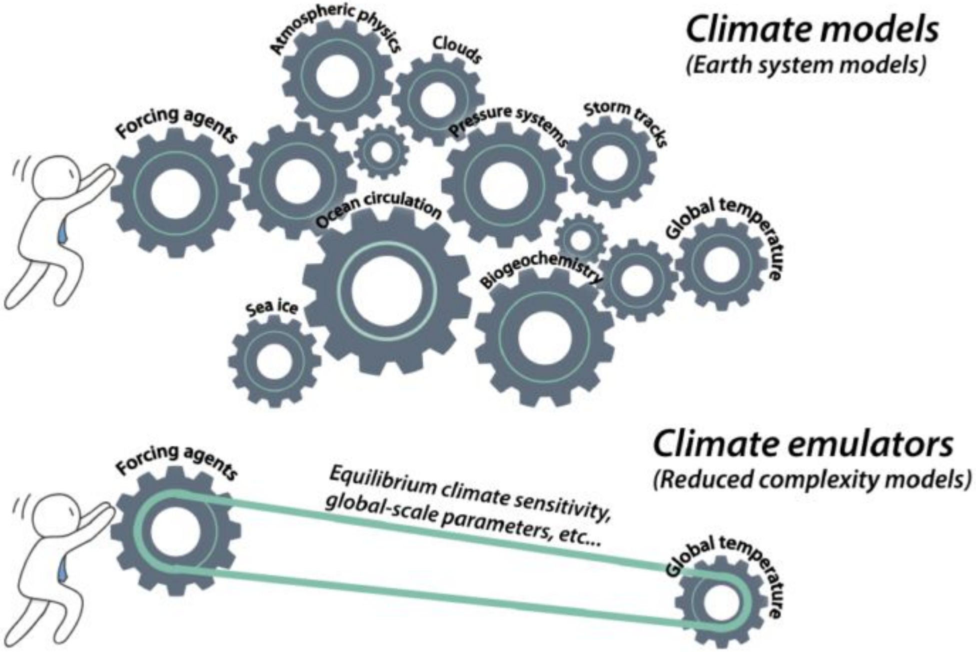
How can machine learning help?

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)

Climate model emulators

Replace the physical simulation with a machine learning model



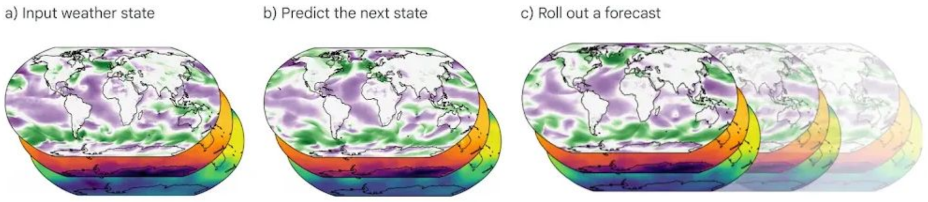
Weather model emulators are performing well as of late

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

14 NOVEMBER 2023

Remi Lam on behalf of the GraphCast team

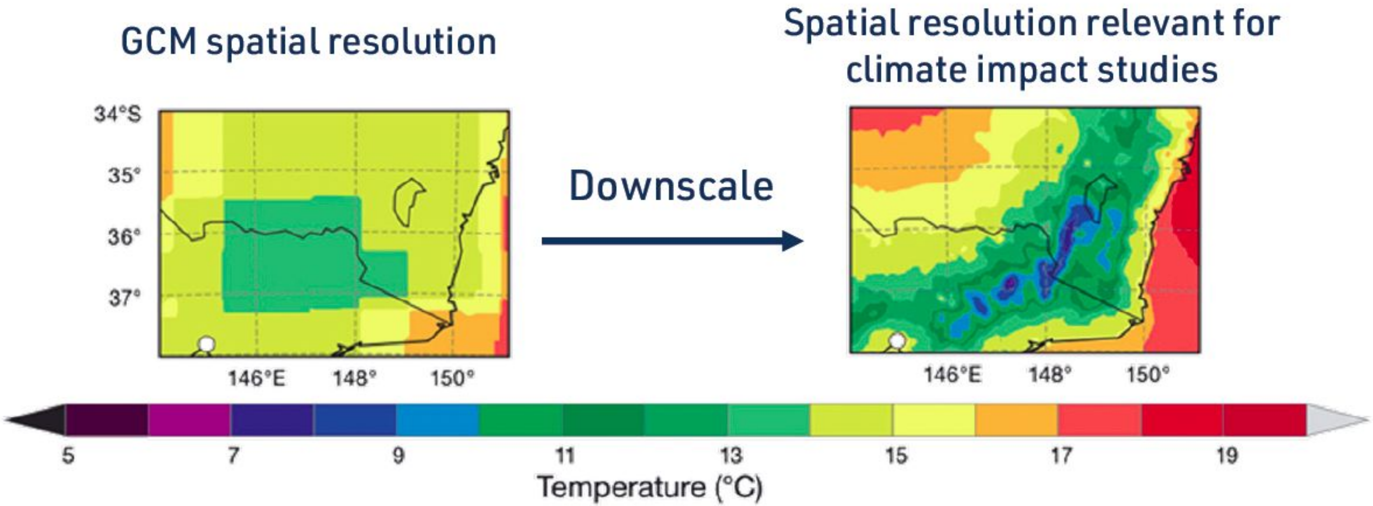
Our state-of-the-art model delivers 10-day weather predictions at unprecedented accuracy in under one minute



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

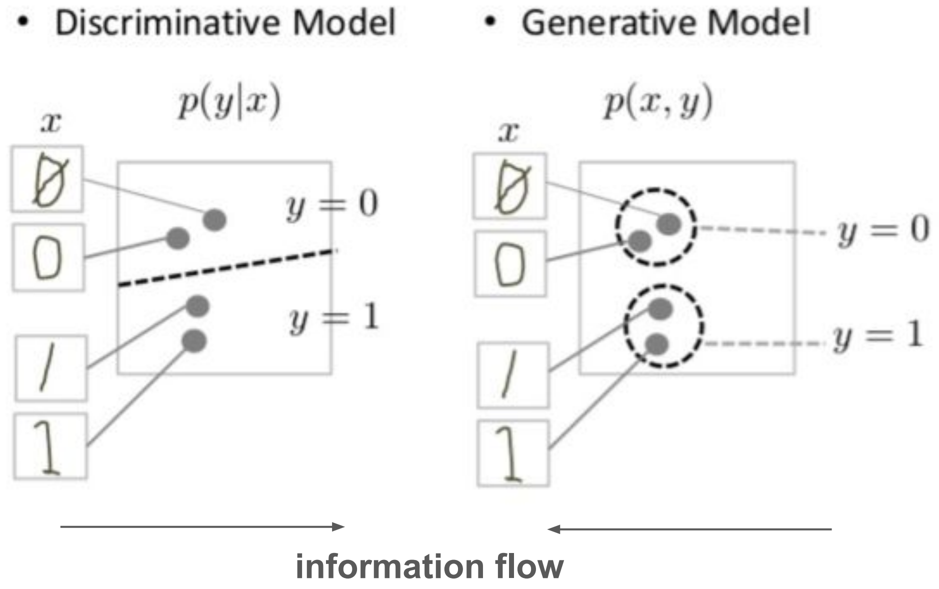


Statistical Downscaling: Learn a model that can project poor resolution information into a higher resolution (“downscaling” = going to a lower spatial scale, “statistical” = learned from data). Also known as “super-resolution”.

This is a “generative” modeling problem

NARCLiM

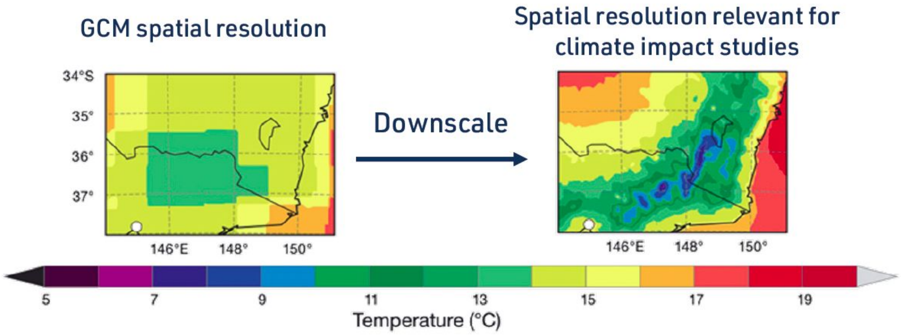
Generative models

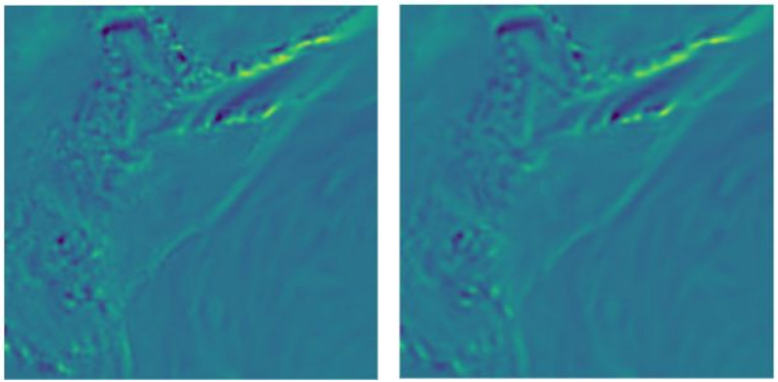


Google

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

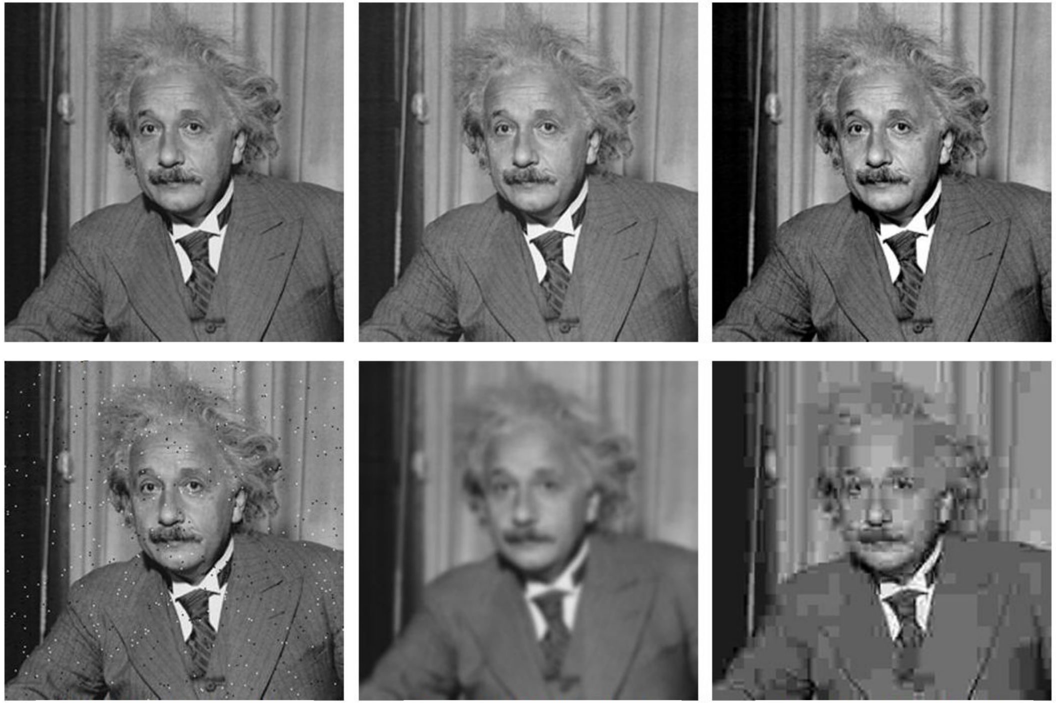




If our generative model is supposed to produce high resolution images of climate variables, we need to be able to answer questions like...

...is the “vorticity spatial fields” (spinning atmospheric wind patterns) on the right a close match to the image on the left?

How do we compare two images?



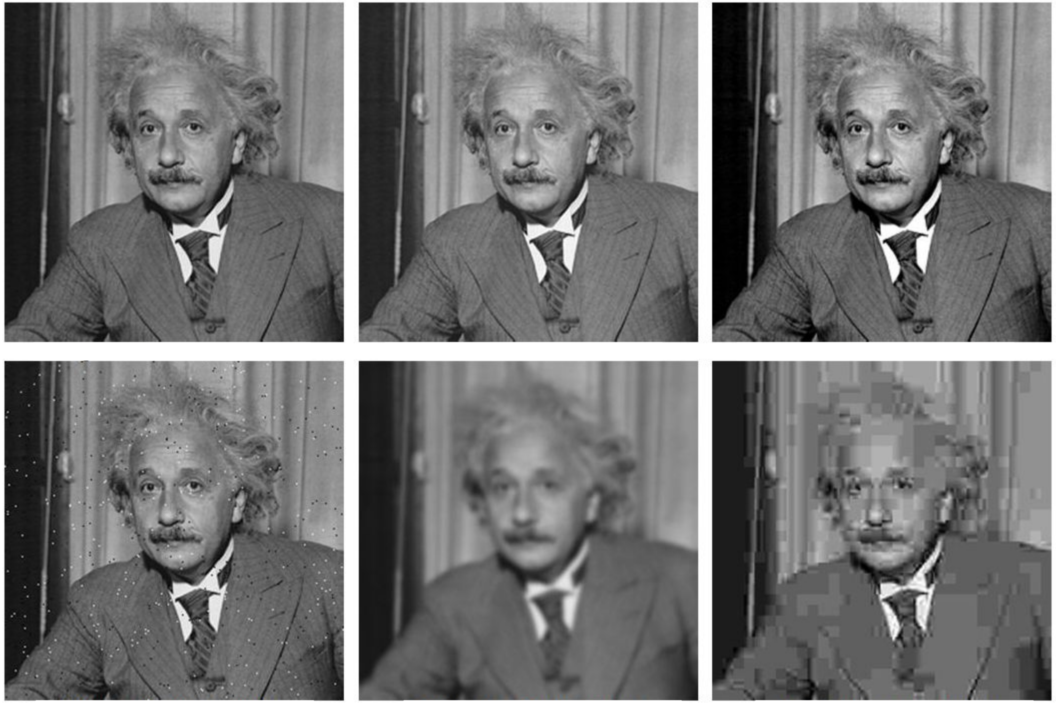
cns.nyu.edu/~lcv/ssim/

How do we compare two images?

We could use 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



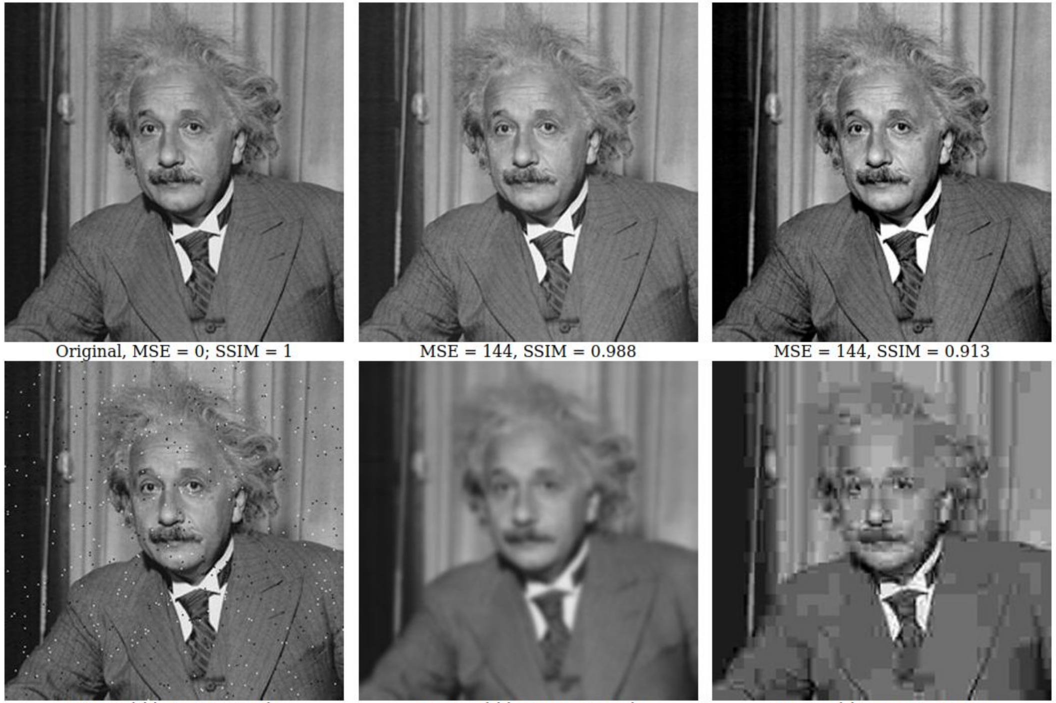
How do we compare two images?

We could use 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

But MSE doesn't necessarily capture what we care about!

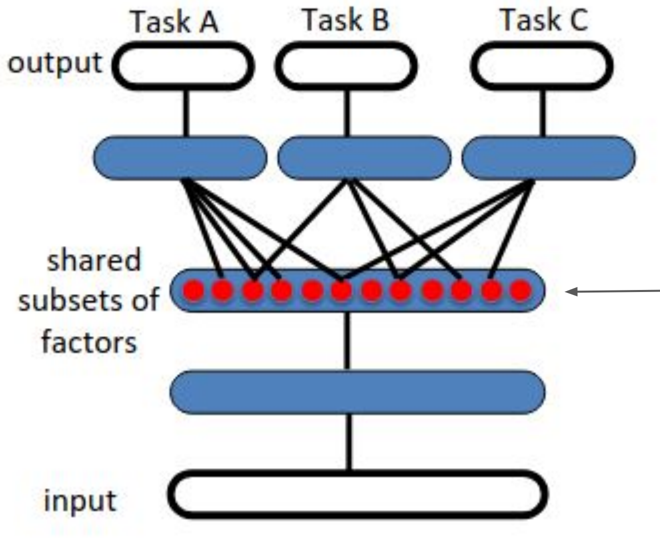


Original, MSE = 0; SSIM = 1
MSE = 144, SSIM = 0.988
MSE = 144, SSIM = 0.913
MSE = 144, SSIM = 0.840
MSE = 144, SSIM = 0.694
MSE = 142, SSIM = 0.662

cns.nyu.edu/~lcv/ssim/

Representation Learning

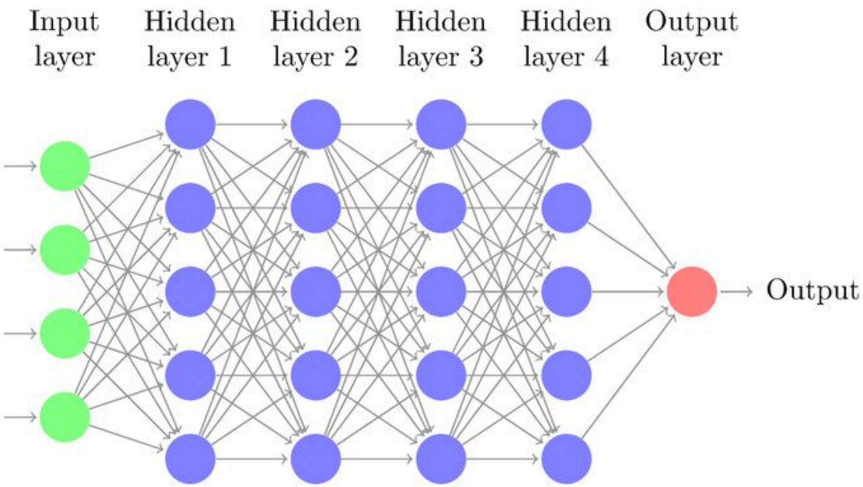
“In machine learning (ML), feature learning or representation learning is a set of techniques that allow a system to automatically discover the representations needed for feature detection or classification from raw data. This replaces manual feature engineering and allows a machine to both learn the features and use them to perform a specific task.” (wikipedia)



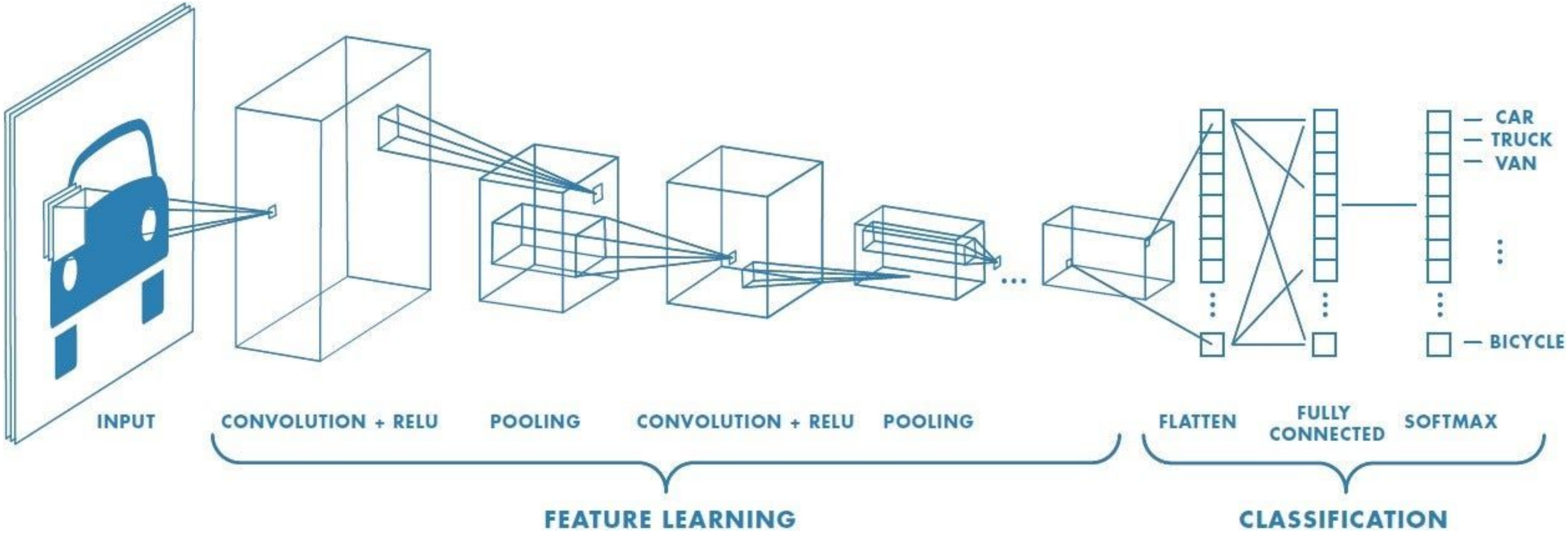
a “good representation” that can be used to perform many tasks

Representation Learning

In trained artificial neural networks, the ‘representation’ usually just refers to the activity of artificial neurons at a certain layer



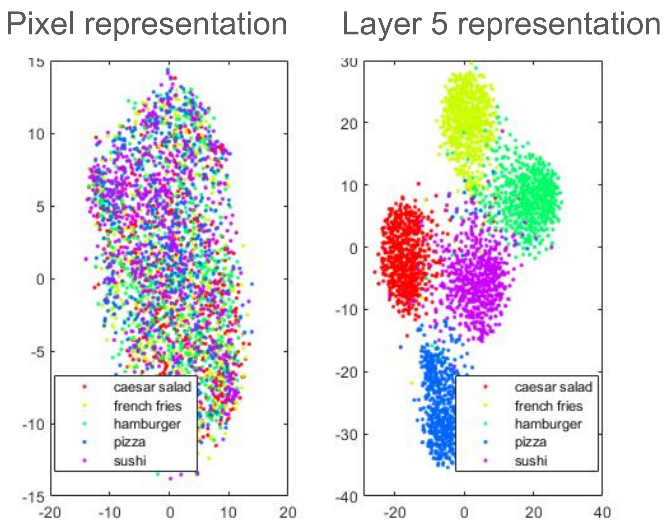
Can CNNs help us compare images?



Convolutional neural networks can learn to represent images in a way that aligns more with both our perceptual experience and how images are used

Can CNNs help us compare images?

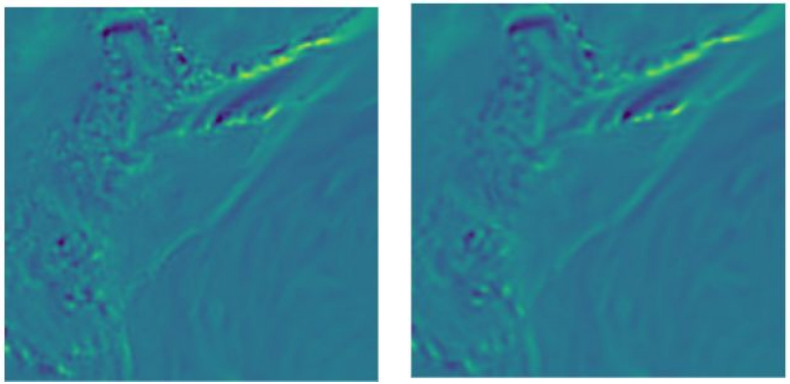
Dimensionality Reduction applied to different layers of a CNN trained on food images:



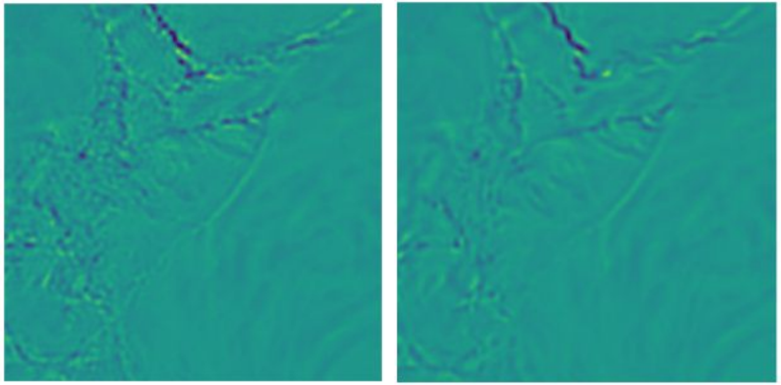
A layer deep into the network better separates different kinds of food

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:

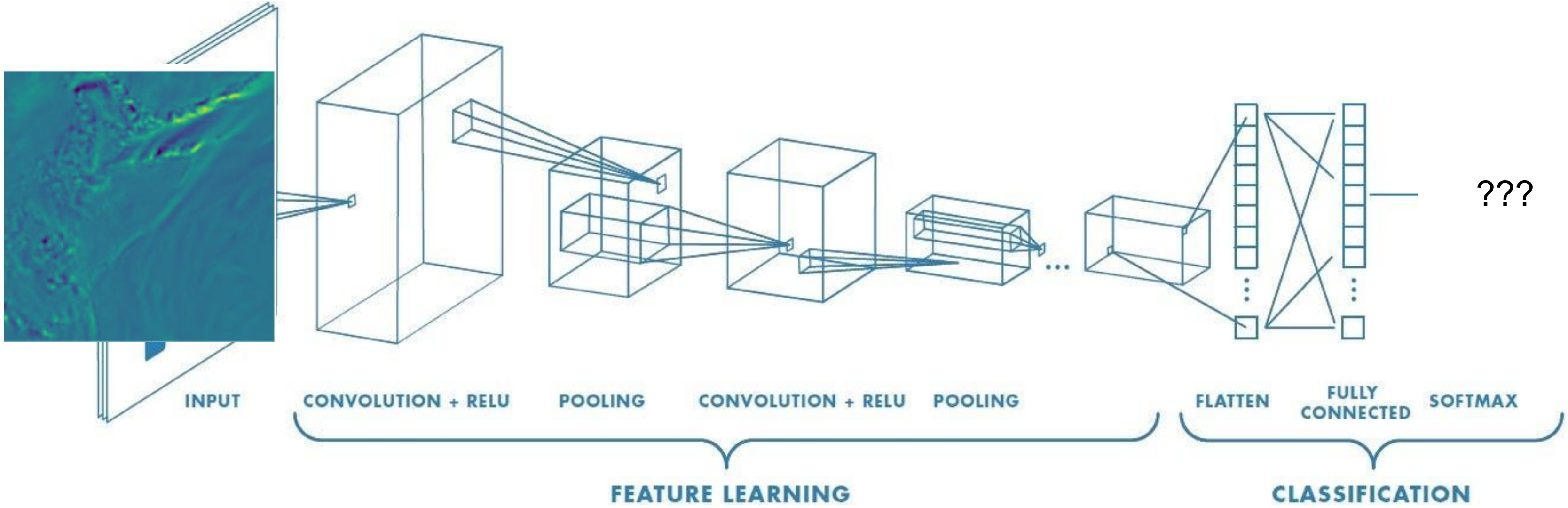


How similar are there two “vorticity spatial fields” (spinning atmospheric wind patterns)?



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

Can we train a CNN that does help?



But what do we train this network to **do**? We don't have labels.
How can we get this model to learn useful representations?

Self-supervised learning doesn't require traditional labels

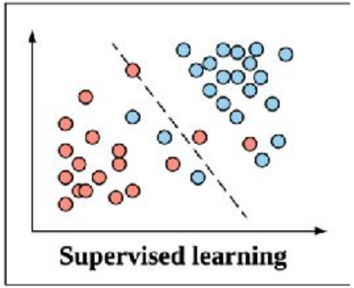
Many different ways to learn representations:

All data points need separate labels

A few data points are labeled

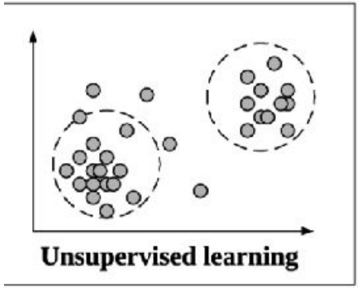
Other aspects of the data are used to create "labels"

No data points are labeled

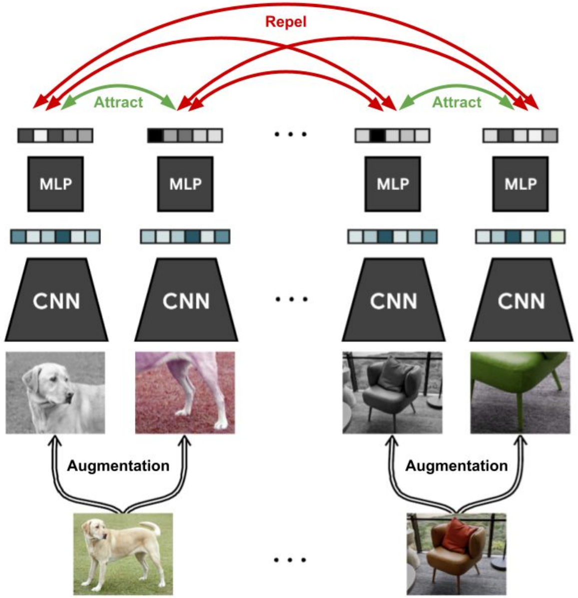


Semi-supervised learning

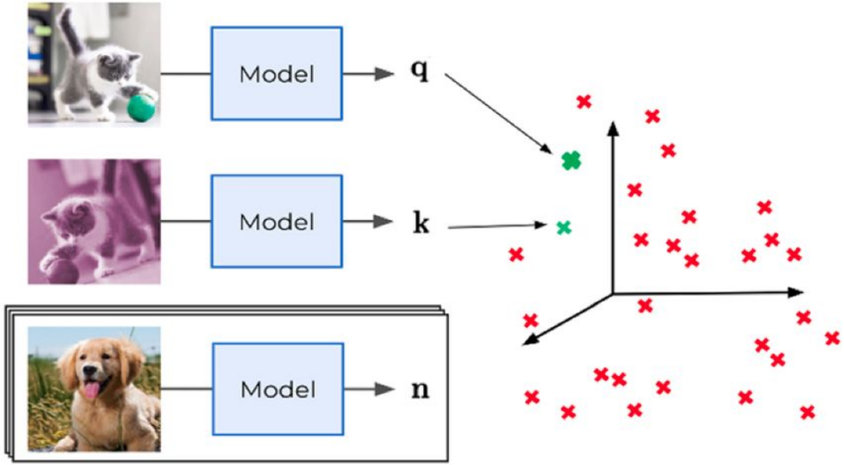
Self-supervised learning



Contrastive learning: type of self-supervised learning



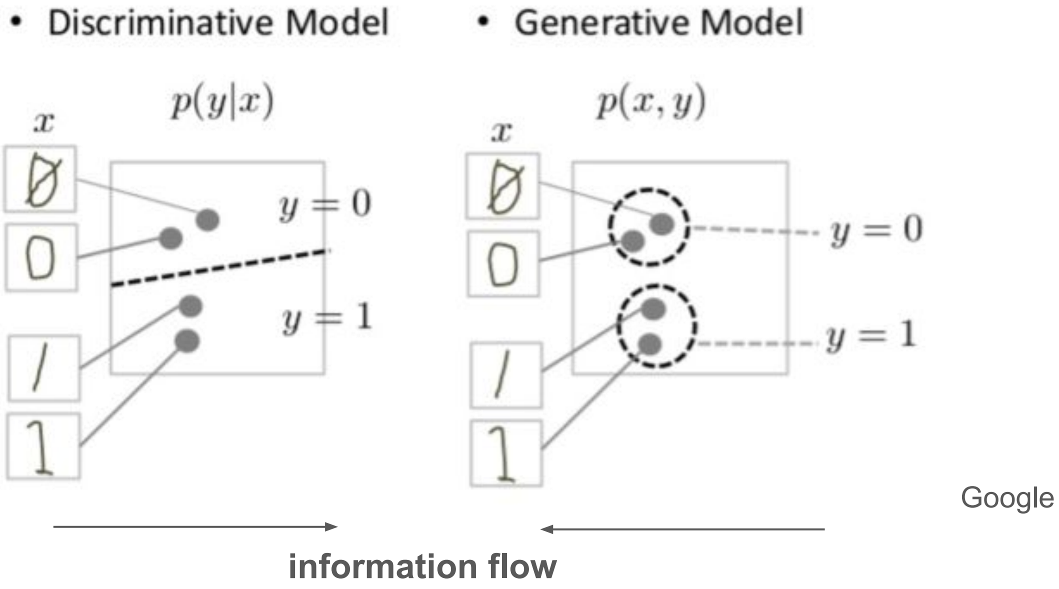
Ting Chen et al



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.

Generative models

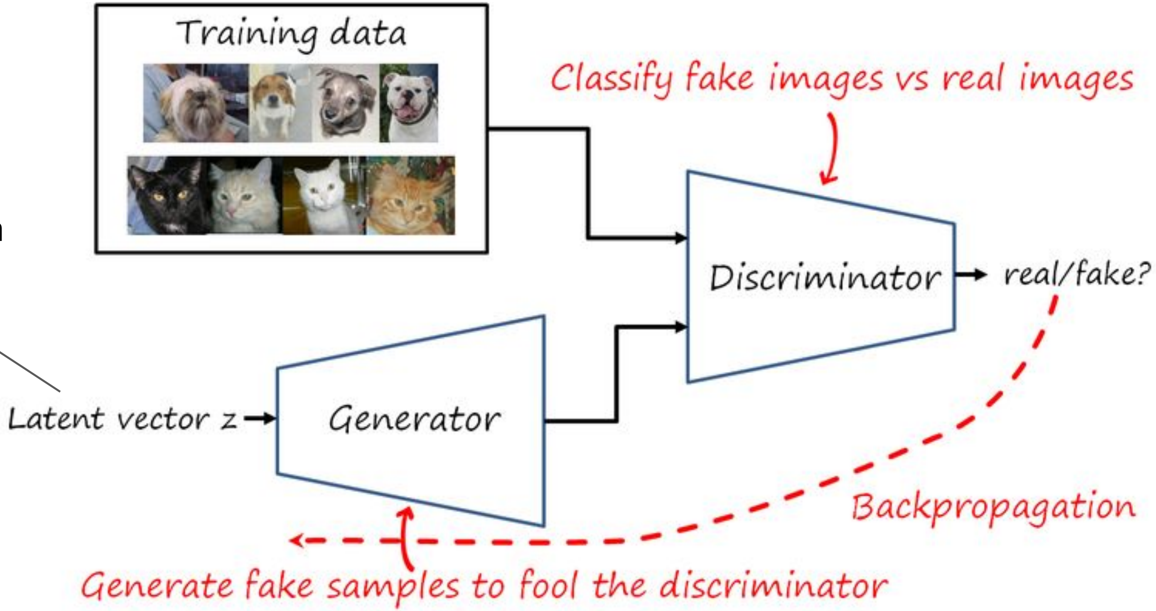


Discriminative: model learns to map a high-dimensional input to a lower dimensional label

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

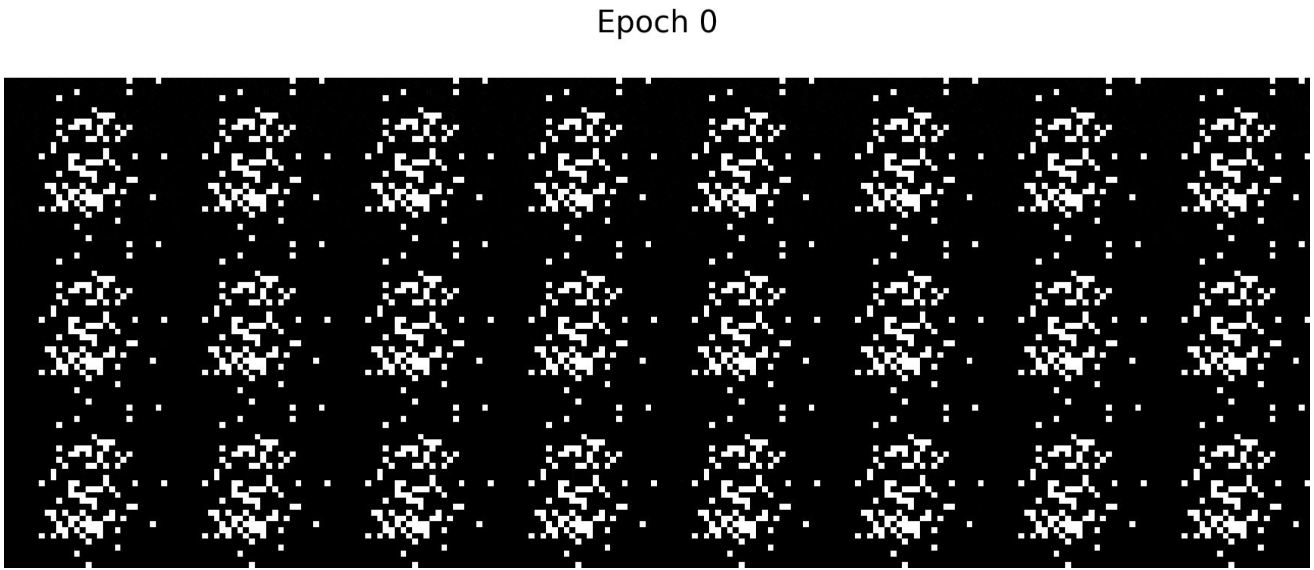
Generative adversarial networks

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



One common architecture 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)



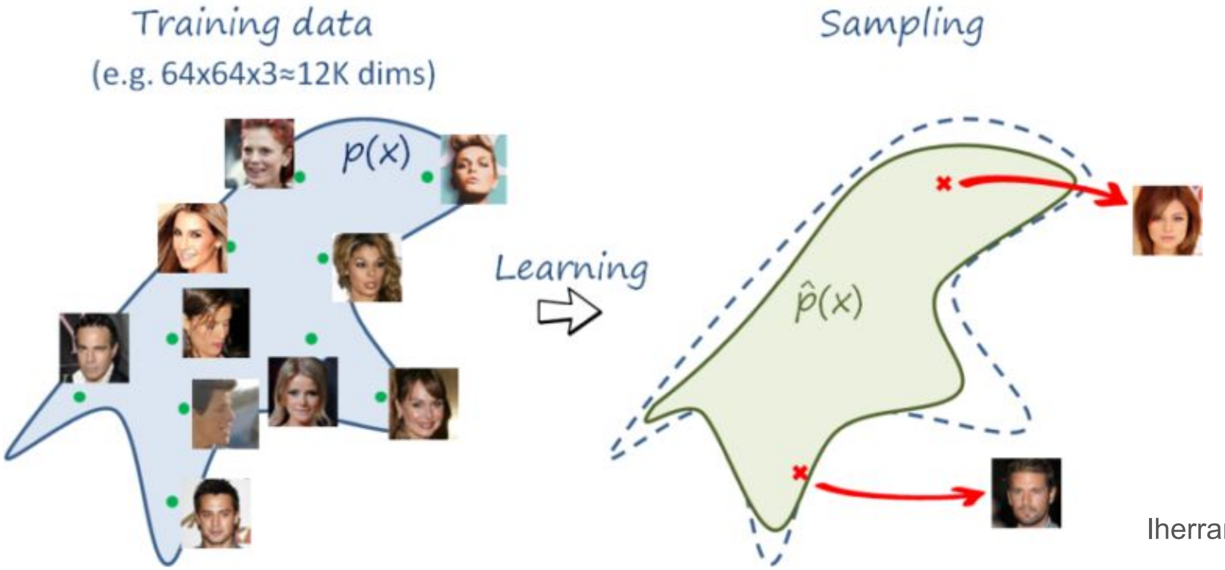
Here, the generator learned the statistics of hand-written digits.

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

thispersondoesnotexist.com

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

Generative models



lherranz.org

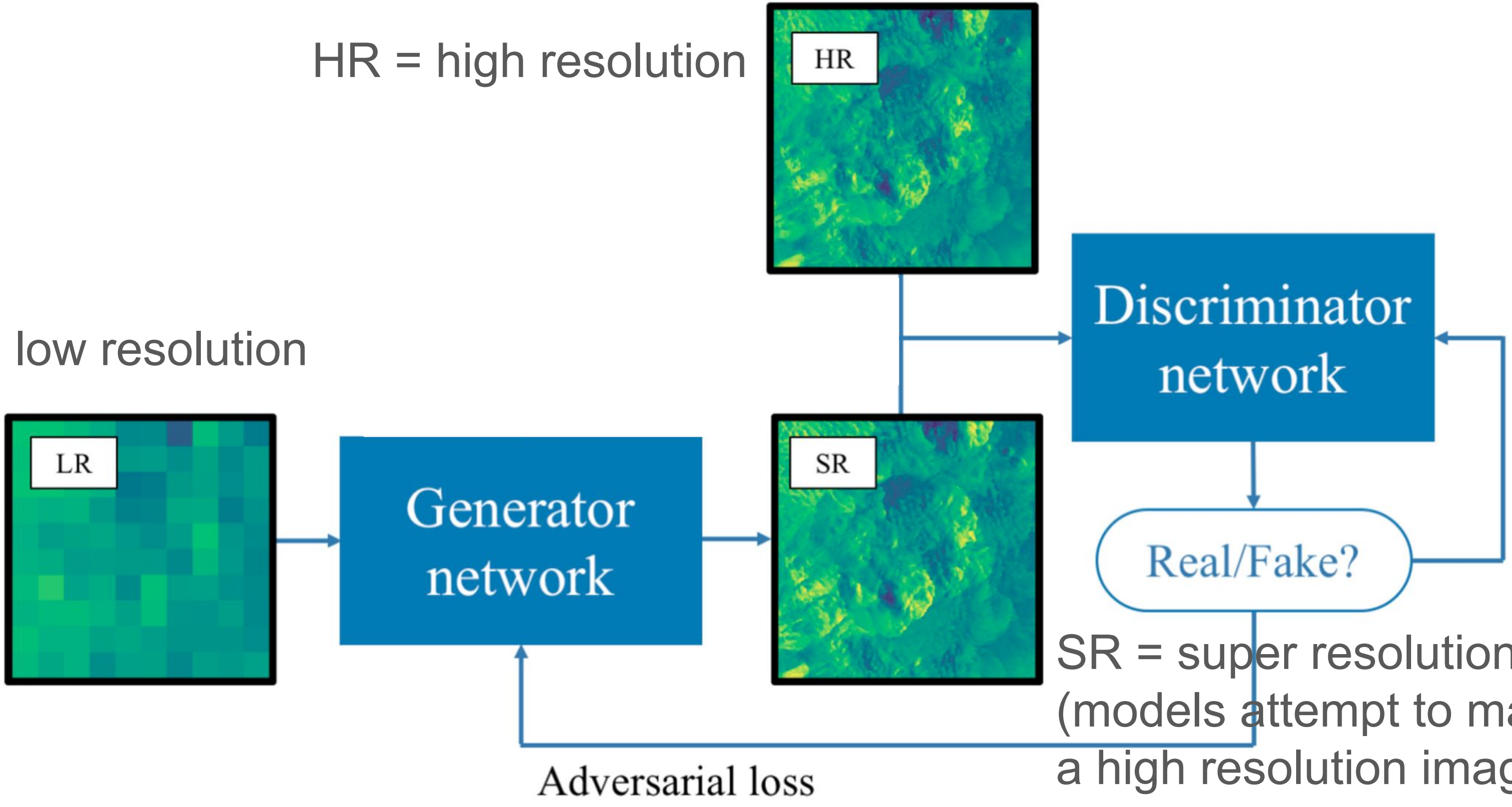
The model learns what patterns of pixel values are statistically likely in face images. It can then generate a new image that has a high probability of being a face.

Because of this, generative models can have many “right answers”

Generative models can be used to solve super resolution

HR = high resolution

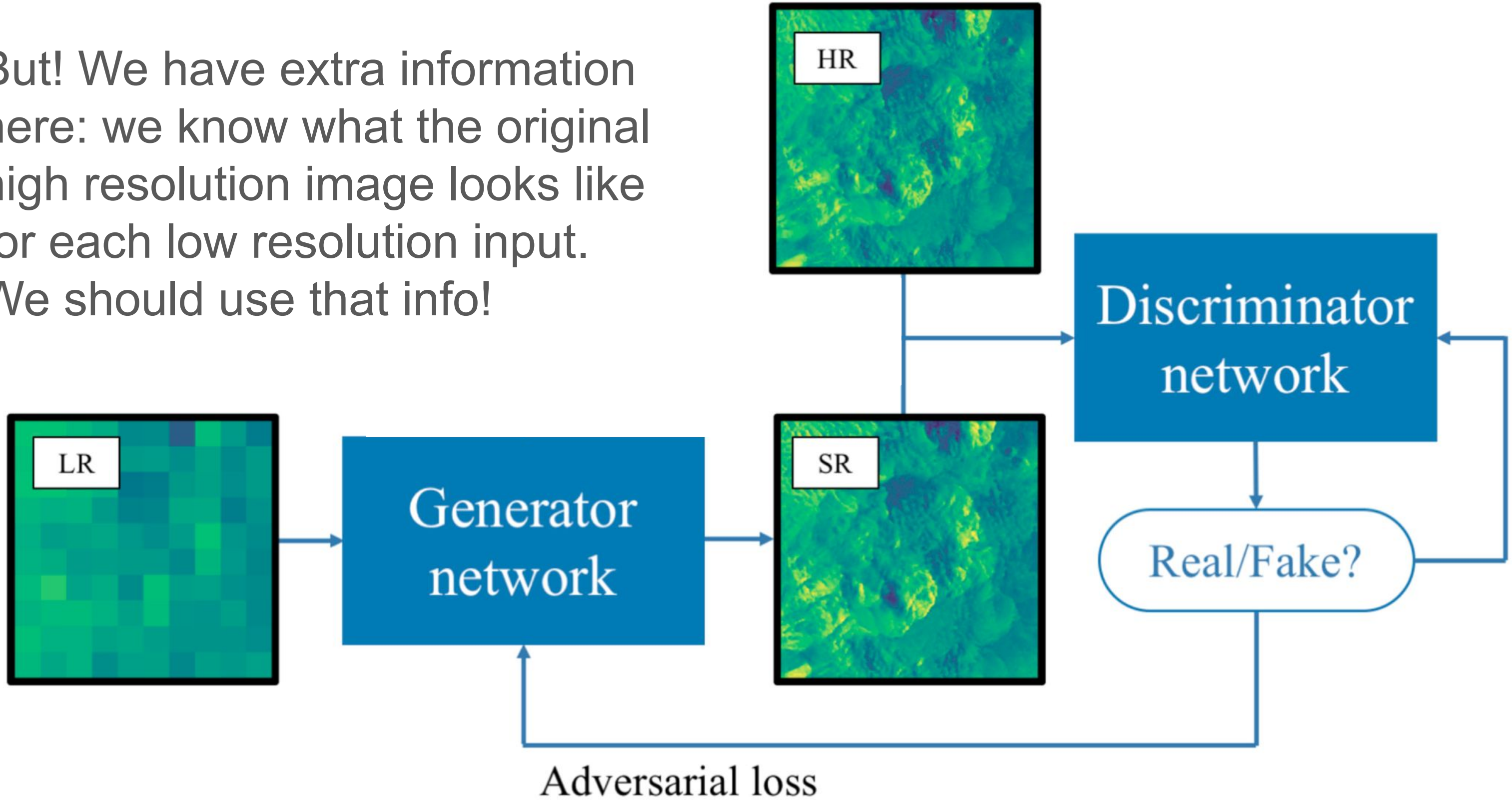
LR = low resolution



SR = super resolution (models attempt to make a high resolution image)

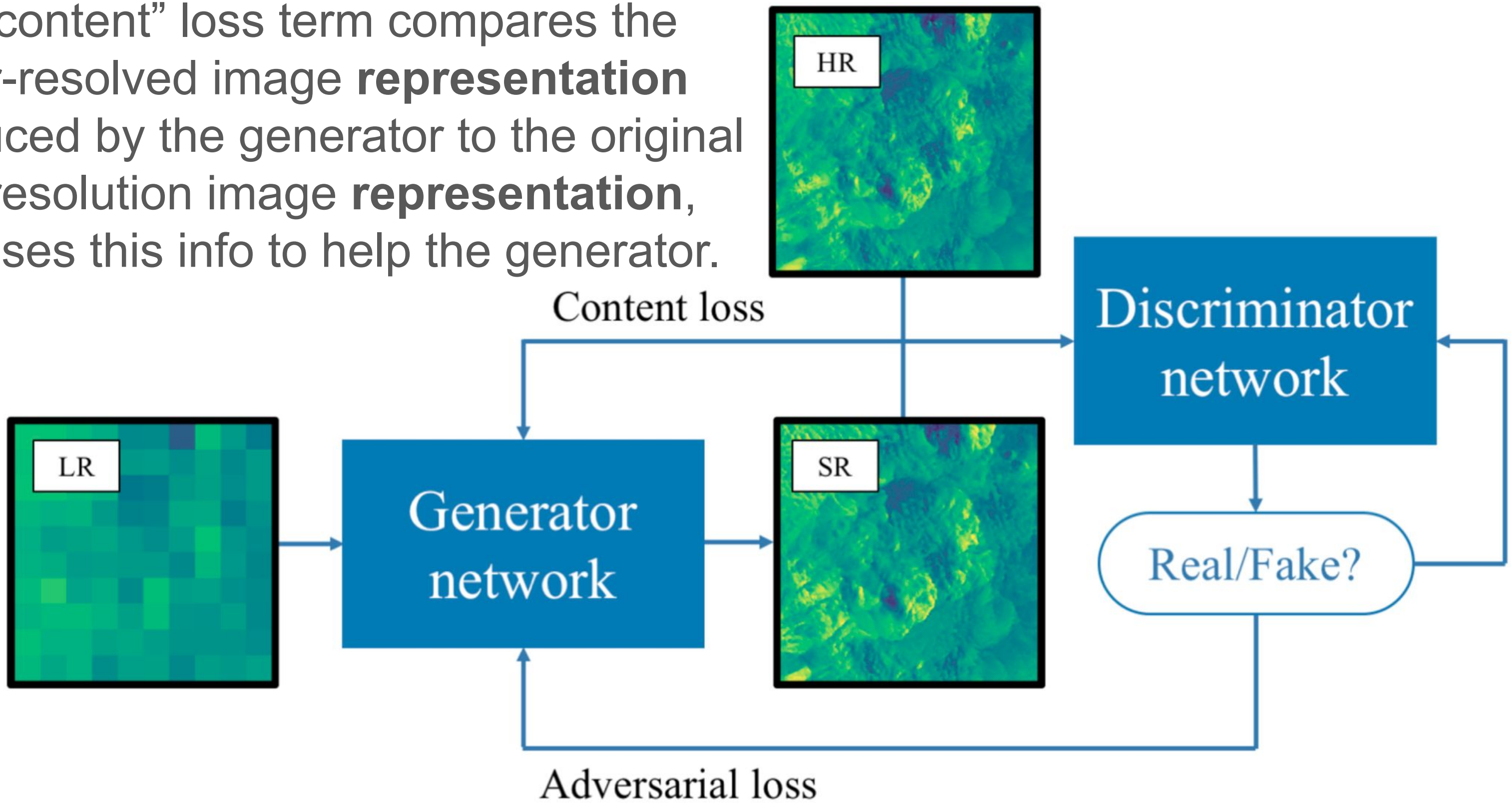
Generative models can be used to solve super resolution

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 **representation** produced by the generator to the original high resolution image **representation**, and uses this info to help the generator.



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