SSP 100 - 10
How Computers See

11 / 12 / 2026

Instructor: Michael Eckmann

Today's Topics

- Questions/Comments?
- Recap on changing basis
- Recap on ways we talked about representing an image for processing and why it is smart
- Revisit Face recognition
  - Current state-of-the-art techniques
  - Some issues with certain datasets

Many examples already of representing an image in some way

- As we learned, for almost all computer vision problems we don't compare raw image pixel values amongst each other.
- Rather we compute something about each image as a representation of the image and use that representation in later processing. One benefit of the representation is that it is a reduction in dimensionality (vs. the original image pixel values).
- Examples:
  - Represent an image as a color histogram
  - Represent a face image as a vector of coefficients of eigenfaces
  - Represent a face image as a vector of 512 values (which is the output of the penultimate layer of a pre-trained CNN)
  - Represent a handwritten digit image as a vector of values highlighting its features

Many examples already of representing an image in some way

- Then we use the reduced dimension representation to compare against other representations in some way
- e.g. compare a 512 valued vector of an unknown face to all the other representations of known faces and find which it is closest to

representation

- Besides dimensionality reduction, the representation contains values highlighting the appropriate features to describe the image well for the problem at hand

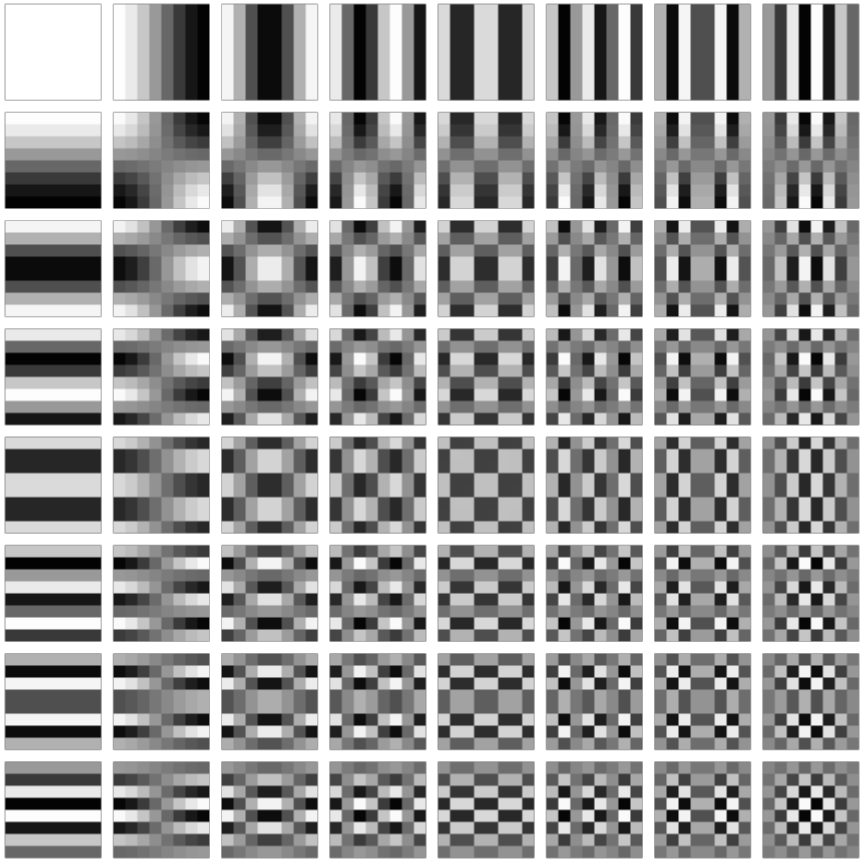
Change of Basis

- Last class was all about changing basis.
- First an appropriate basis needs to be selected
  - Then transforming into that basis computes a set of coefficients of each of the basis vectors/images
  - Then we can further reduce the data by either
    - keeping only the coefficients with the highest absolute value and discard any low or 0 coefficients
    - Or keeping certain coefficients over others based on something else besides value
  - This is one way to provide dimensionality reduction as well as good descriptive power

Change of Basis

- Two good examples highlight the appropriateness of basis for the particular task
  - JPEG use of 8x8 Discrete Cosine Transform basis images
  - Eigenfaces as a basis to describe faces well.
  - What's the task for each of these situations?
  - How are these appropriate for the task?

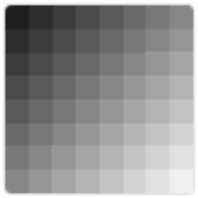
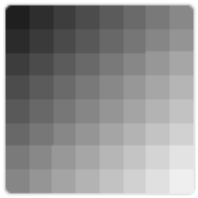
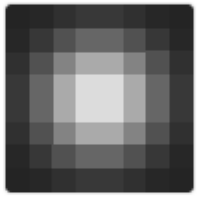
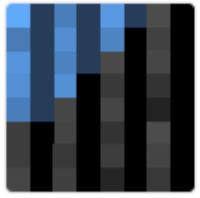
Use of DCT in JPEG compression



Michael Eckmann - Skidmore College - SSP 100 - Fall 2026

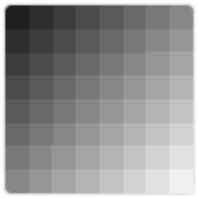
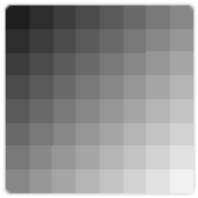
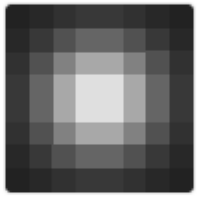
Coefficients kept 20 / 64

Which ones | Lowest frequencies (zig-zag) | Largest magnitude | 3.2× fewer numbers

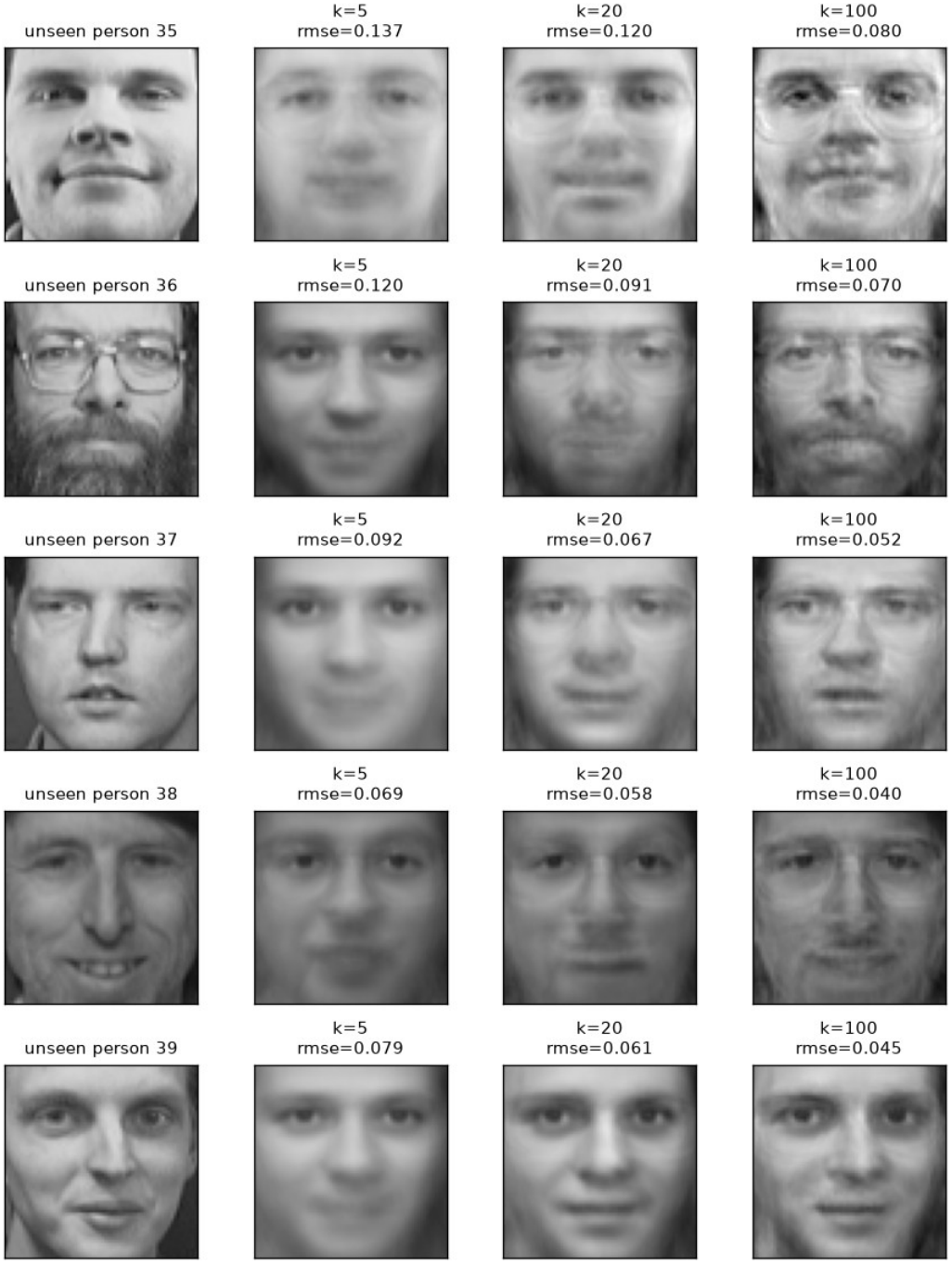
	Original	DCT coefficients	Reconstruction	Quality
Smooth gradient diagonal ramp				Detail kept **99.9%** RMSE **1.1** · PSNR **47.1 dB** Needs **3** coeffs for 95%
Soft blob Gaussian spot				Detail kept **99.9%** RMSE **1.3** · PSNR **46.0 dB** Needs **4** coeffs for 95%
Hard edge vertical step				Detail kept **96.8%** RMSE **15.3** · PSNR **24.4 dB** Needs **4** coeffs for 95%
Letter A binary glyph				Detail kept **32.9%** RMSE **76.1** · PSNR **10.5 dB** Needs **21** coeffs for 95%
Checkerboard highest frequency				Detail kept **0.4%** RMSE **92.3** · PSNR **8.8 dB** Needs **8** coeffs for 95%

Coefficients kept 20 / 64

Which ones | Lowest frequencies (zig-zag) | Largest magnitude | 3.2× fewer numbers

	Original	DCT coefficients	Reconstruction	Quality
Smooth gradient diagonal ramp				Detail kept **100.0%** RMSE **0.0** · PSNR **∞** Needs **3** coeffs for 95%
Soft blob Gaussian spot				Detail kept **100.0%** RMSE **0.0** · PSNR **∞** Needs **4** coeffs for 95%
Hard edge vertical step				Detail kept **100.0%** RMSE **0.0** · PSNR **∞** Needs **4** coeffs for 95%
Letter A binary glyph				Detail kept **94.4%** RMSE **21.2** · PSNR **21.6 dB** Needs **21** coeffs for 95%
Checkerboard highest frequency				Detail kept **100.0%** RMSE **0.0** · PSNR **∞** Needs **8** coeffs for 95%

Faces the model never saw: reconstructed well



Change of Basis

How do these achieve dimensionality reduction?

- Examples showed for the DCT if we keep the 20 highest values of the 64 coefficients we get a really good descriptor for many situations. The representation is 20 values --- reduced 64 values to 20
- For face examples, those face images were 64x64 grayscale = 4096 values and the right most column were the reconstructed images from just the top 100 coefficients --- reduced 4096 values to 100

Last 15 or so years

- Convolutional Neural Networks trained in a supervised manner.
- Images of faces with known identities used for training --- that is supervised learning because the data (images of faces) and labels (known identities) are provided.
- Number of output nodes in the network would be the number of identities the network can detect.

Some datasets (table generated by Claude)

Current mainstream training sets (real, web-scraped)

Dataset	Identities	Images	Imgs/ID	Derived from	Role today
MS1MV2	85K	5.8M	~68	MS-Celeb-1M, cleaned	Long-time standard baseline; most papers still report results on it
MS1MV3	93K	5.2M	~56	MS-Celeb-1M, re-cleaned	Cleaner drop-in replacement for MS1MV2
Glint360K	360K	17M	~47	MS1M + Celeb-500K, merged and cleaned	InsightFace's large set (2021), popular for big-backbone training
WebFace4M	200K	4.2M	~21	10% subset of WebFace42M	Very widely used now; same scale as MS1M but trains better models
WebFace12M	600K	12M	~20	30% subset of WebFace42M	Mid-scale option; AdaFace releases checkpoints on it
WebFace42M	2M	42M	~21	WebFace260M, auto-cleaned (CAST)	Largest public cleaned training set
WebFace260M	4M	260M	~65	Raw web scrape	Noisy source pool; not trained on directly

VGGFace2

- Retracted due to ethical concerns of lack of consent and privacy concerns mainly.
- As with any pretrained model, there are also concerns about bias and under or over representation of certain groups which could lead to more errors with recognizing people of those underrepresented groups.

MS-Celeb-1M

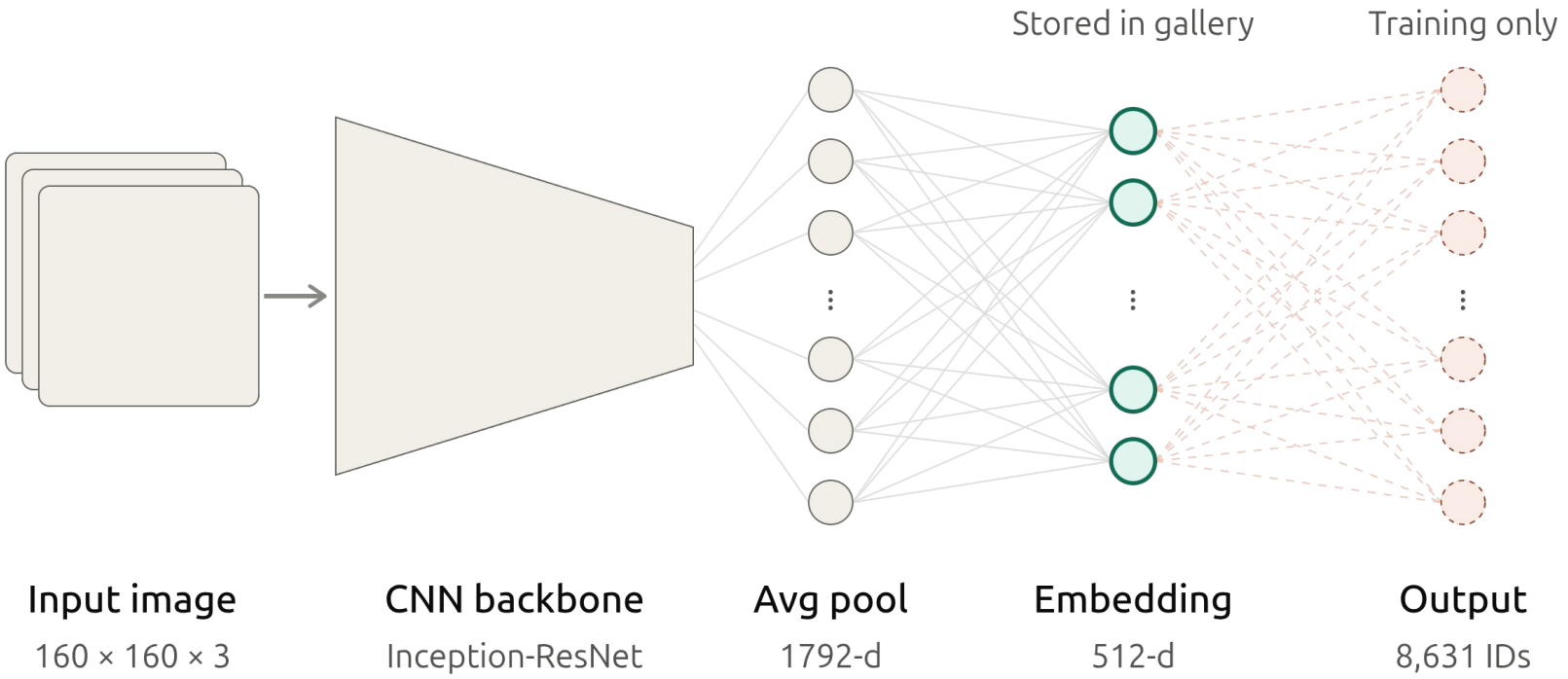
- MS-Celeb-1M by Microsoft also was taken down after complaints that it included other public figures that are not celebrities.

- **https://exposing.ai/msceleb/**

You too can do face recognition!

- How?
- Publicly available pre-trained models exist
- Take the model's trained weights up to but not including the last layer. The output of the penultimate layer is called an embedding.

You too can do face recognition!



You too can do face recognition!

- Take a pretrained model that converts images of faces into an embedding (exclude the weights after the penultimate layer and the output layer).
- Run your dataset through it to compute an embedding per identity.
- Then send a test face image through to compute the embedding of that face and compare the embedding to your database of embeddings.

Enrollment

- Enroll your dataset
  - Have a small number of reference images per person
  - For each person:
    - Compute the embedding of each reference image of a person (by running each through the pretrained model)
    - Store a normalized average embedding per person

Probe & Match attempt

- An image of a person is captured (the probe)
  - Detect the face within the image
  - Do some processing to the face portion of the image (e.g. straighten it) and make it the correct dimensions and number of pixels required (e.g. 160x160)
  - Compute the embedding (by running it through the pretrained model)
  - Compare this embedding to the known embeddings for recognition
  - Return the identity of the best match

Matching

- Think about potential answers a probe could give.

Matching

- Think about potential answers a probe could give.
- Face will be recognized as someone in your database, but what if the person detected isn’t in your database?

Matching

- Think about potential answers a probe could give.
- Face will be recognized as someone in your database, but what if the person detected isn't in your database?
- An unknown person should not be recognized incorrectly as a person in your database.
- This is the problem with closed set recognition --- closed set recognition always gives an answer of a known identity.
- Open set recognition should allow an additional output label of "Unknown".

Matching

- For open set recognition to allow an additional output label of “Unknown” what is typically done is to first get the highest matching score for a person within the database.
- If this highest matching score is above some preset threshold, then answer that person’s identity.
- If the highest matching score is below the threshold, output “Unknown.”
- Other techniques can be done when that simple way doesn’t work well.

Demo

- Let me run that demo program I have already and talk about what its doing

Matching

- For a face recognition system you'd want to test it and record a variety of different errors:
- FPIR – False positive identification rate
  - when non-enrolled people are identified as an enrolled person
- FNIR – False negative identification rate
  - What do you think a false negative means?

Matching

- For a face recognition system you'd want to test it and record a variety of different errors:
- FPIR – False positive identification rate
  - when non-enrolled people are identified as an enrolled person
- FNIR – False negative identification rate
  - when enrolled people are not identified correctly

Discussion

- I'd like you to discuss in groups your thoughts on:
- The ethics of scraping images that are on the internet and using them to train a face recognition system
  - Is it the same for everyone or is it more o.k. for some groups of people e.g. famous vs. non-famous, celebrities vs. others, politicians, regular people posting images on facebook, etc.
- Suppose the dataset is taken down, but the model trained on scraped images on the internet lives on and is used in the manner I described earlier --- to be able to create embeddings of a different group of people (like say employees at a company for access to the building for example)
  - What issues do you feel there are in using the model (without the final layer) and why?

Discussion

I'd like someone in each group to take notes thoughts that came up (without identifying each individual)

- For each of the two questions I'd like to know how many side more towards ethical vs. non-ethical in your group.
- Also I'd like to hear reasons given for either for or against.
- Also if your group is homogenous in its votes toward ethical / non-ethical, I'd appreciate if the group discussed at least reasons why they think different people would think otherwise and note those. That is, I'm asking for someone to play devil's advocate during the discuss of the ethics.