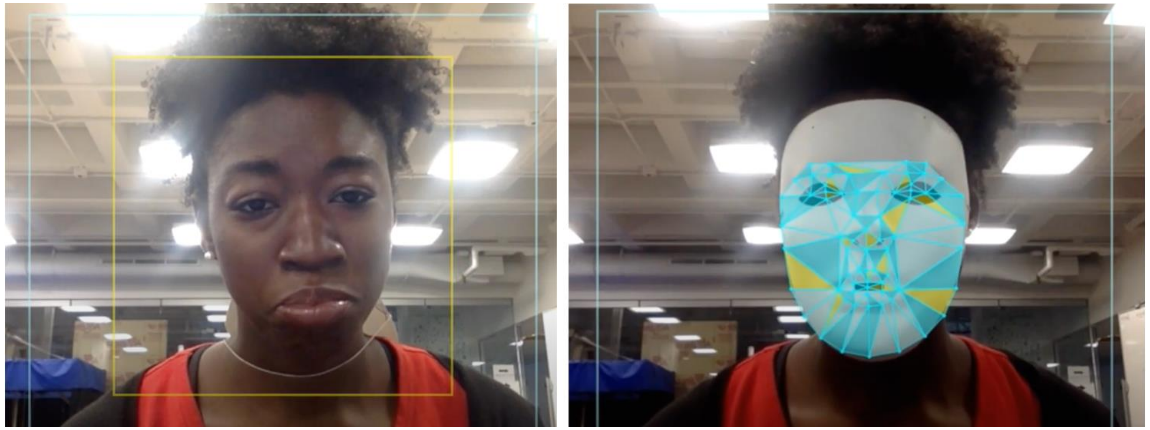
Facing Up To Bias

Steve Teig
Perceive

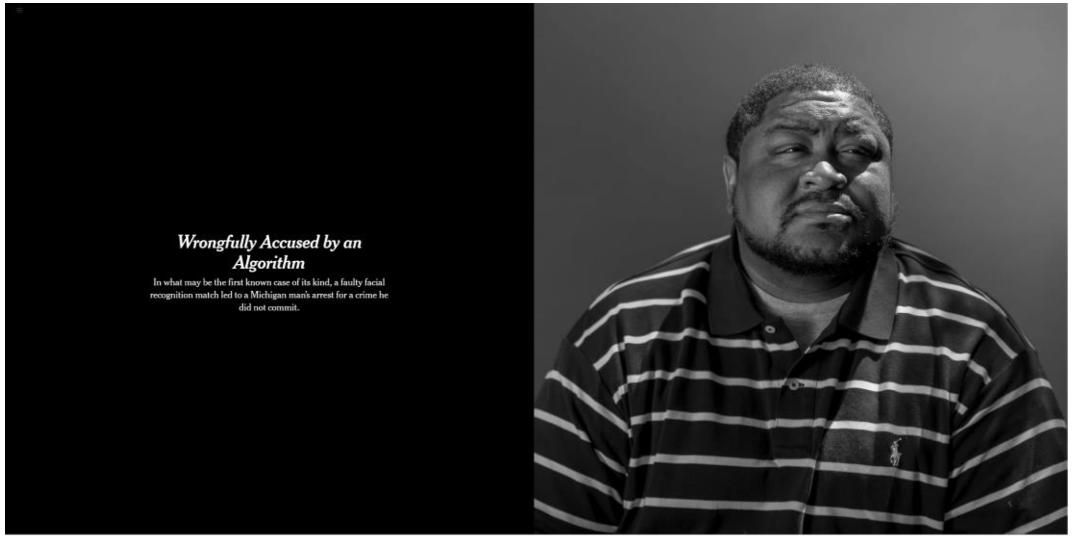
2021 embedded VISION summit VIRTUAL | MAY 25-27

Perceive™

The concerning state of face recognition (FR)



Khan Academy

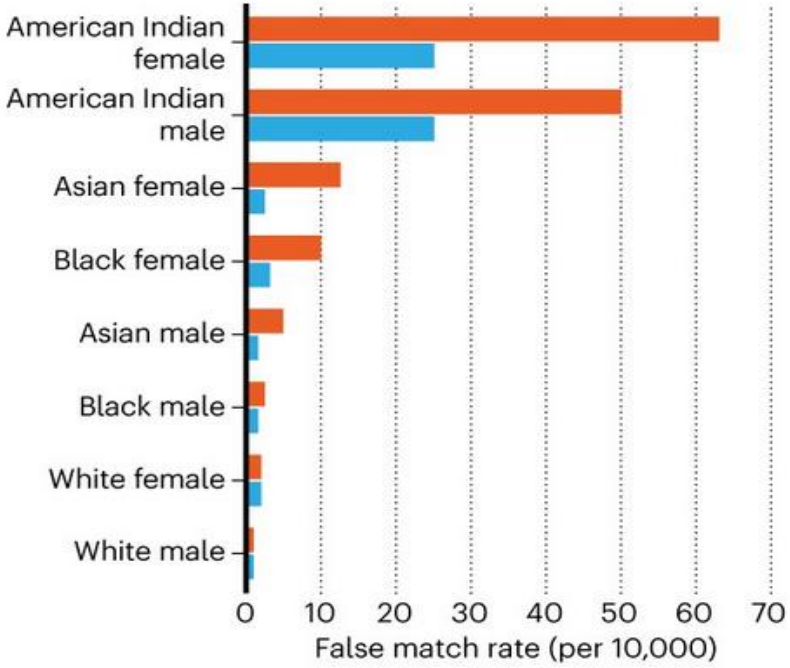


New York Times

MISTAKEN IDENTITY

A 2019 review of facial-recognition algorithms shows the chance of false positives* — incorrectly finding matches between two faces — when comparing high-quality US mugshots of different people of the same gender and race†. The rate is highest for female faces of people of colour, but differs across algorithms (shown in two examples).

- UK academic algorithm
- Chinese commercial algorithm



*Algorithm's confidence threshold for a 'match' was set so as to ensure the false-positive rate for white males was 1 per 10,000; others used same threshold. †Ethnicities as described in ref. 5.

©nature

Perceive™

Backlash and emerging legislation

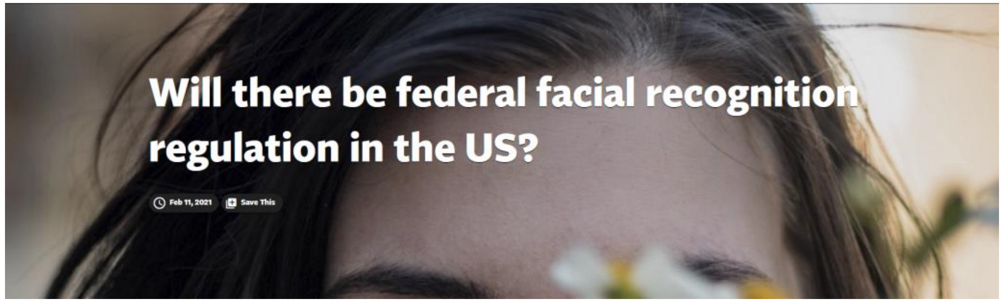
NEWS FEATURE · 18 NOVEMBER 2020

Resisting the rise of facial recognition

Growing use of surveillance technology has prompted calls for bans and stricter regulation.

Antoaneta Roussi

Cameras watch over Belgrade's Republic Square. Credit: Vladimir Zivojinovic for *Nature*

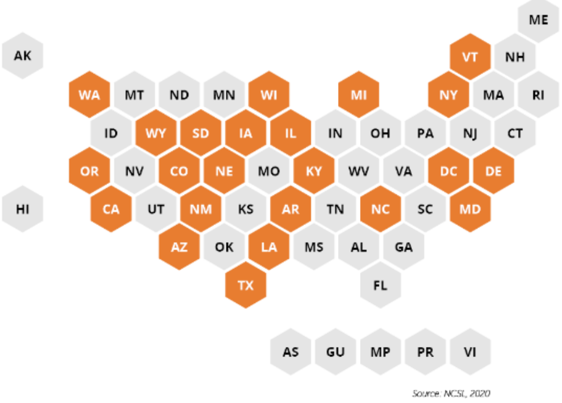


Artificial intelligence + Add to myFT

EU backs away from call for blanket ban on facial recognition tech

New draft of AI paper drops suggestion of 5-year moratorium on surveillance technology

Brussels' latest draft AI paper puts the onus on individual member states to assess how and when they wish to permit the use of facial recognition © AFP via Getty Images



Expanding State Privacy Legislation

Security breach notification laws in at least 22 states and the District of Columbia expressly include biometric information in their definitions of covered personal information.

Expressly include biometric information

Discrimination is pervasive…

2021 embedded VISION summit

~~Sexism~~

~~Sinophobia~~

~~Racism~~

~~Hispanophobia~~

~~Xenophobia~~

~~Ageism~~

Perceive™

Discrimination is pervasive… <u>but not the whole story</u>

- Training a neural network (typically) minimizes a *loss function*
- Near-universal loss function: expected value – i.e., the *average* – of the error
  - E.g., cross-entropy $H(p,q) = -E_p[\log q]$ = average over p of $-\log(q)$

- Suppose our FR training set has 10,000 white faces and 100 black faces
- $Error_W$ = average error on white faces; $Error_B$ = average error on black faces
- Total error is proportional to $10{,}000 * Error_W + 100 * Error_B$

- Yup. Average error penalizes errors on white faces <u>100x</u> as much as errors on black faces!

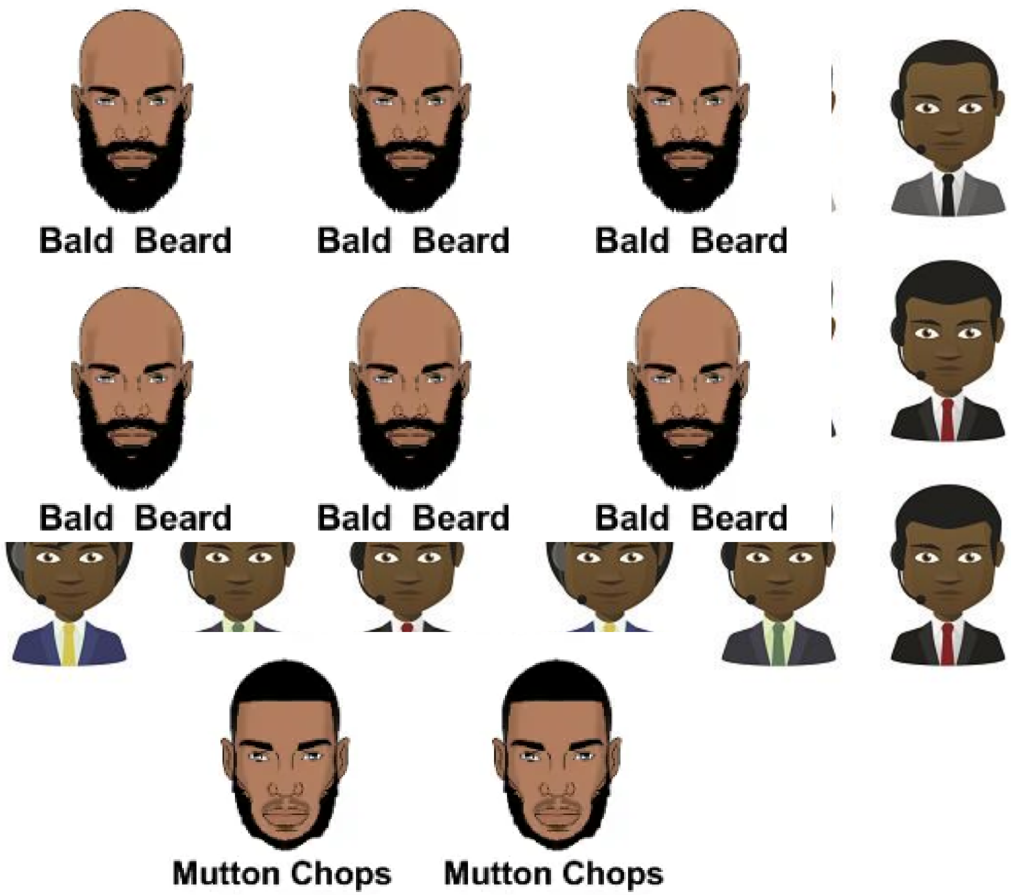
Perceive™

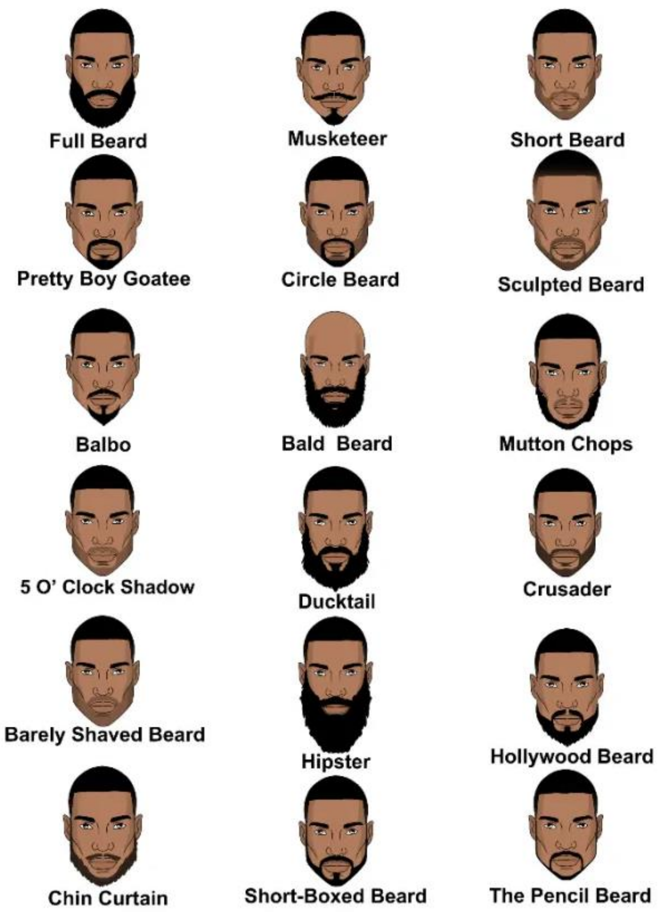
Of course, the trained model does better on white faces!

- Total error ∝ 100 * Error_W + 1 * Error_B
- Average error penalizes errors on white faces 100x as much as errors on black faces!

- Model compression makes this problem even worse
- Quantize the network, sparsify the network, etc.
- If the training network must jettison some information…

Why “balancing” the dataset won’t fix this





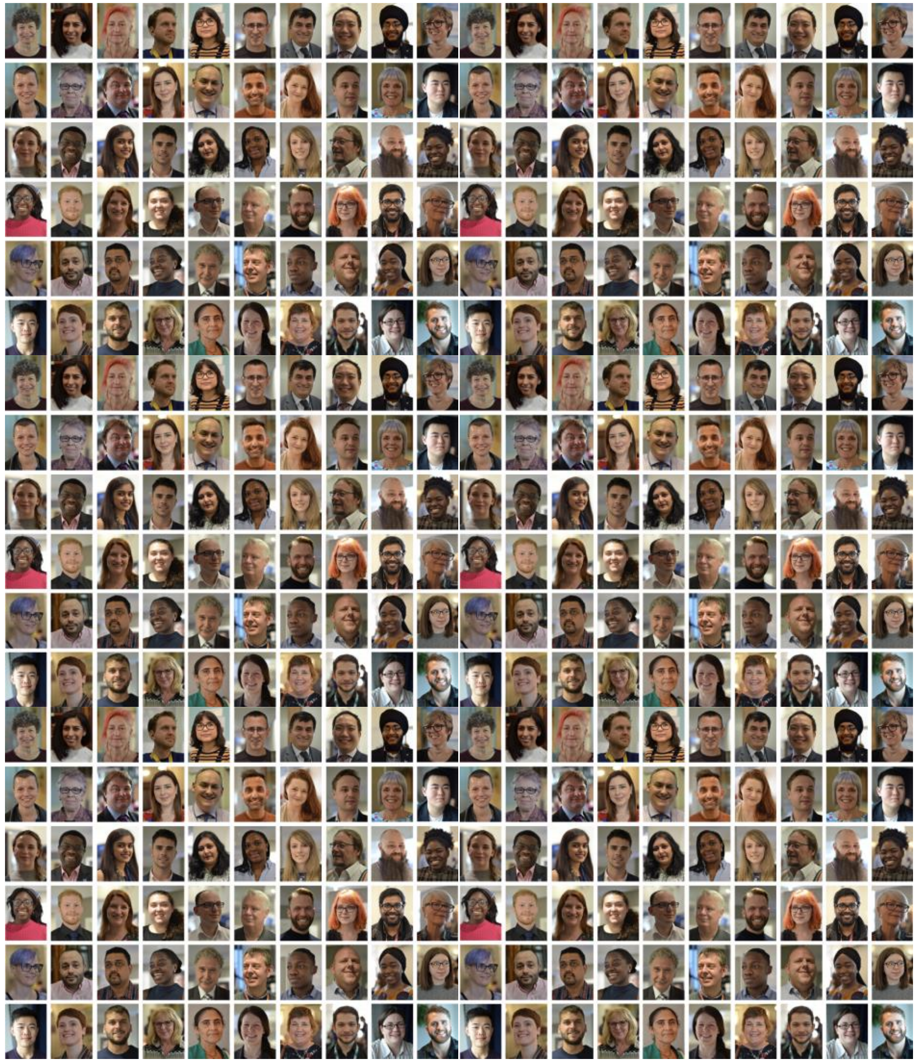
Perceive

For experts: why (naïve) GANs won't fix this either

- GAN: Generative Adversarial Network
  - Generates synthetic data points that are hard to distinguish from real data points
- Can't we use GANs to add more representative, interesting examples to the dataset?
- Yes, but…
- Mainstream GANs optimize only "datum looks as though from the original dataset"
- What if synthetic, clean-shaven faces are easier to generate than bearded ones?
- What if white faces are easier to generate than black ones?
- More bias ☹

How much influence should one image have?

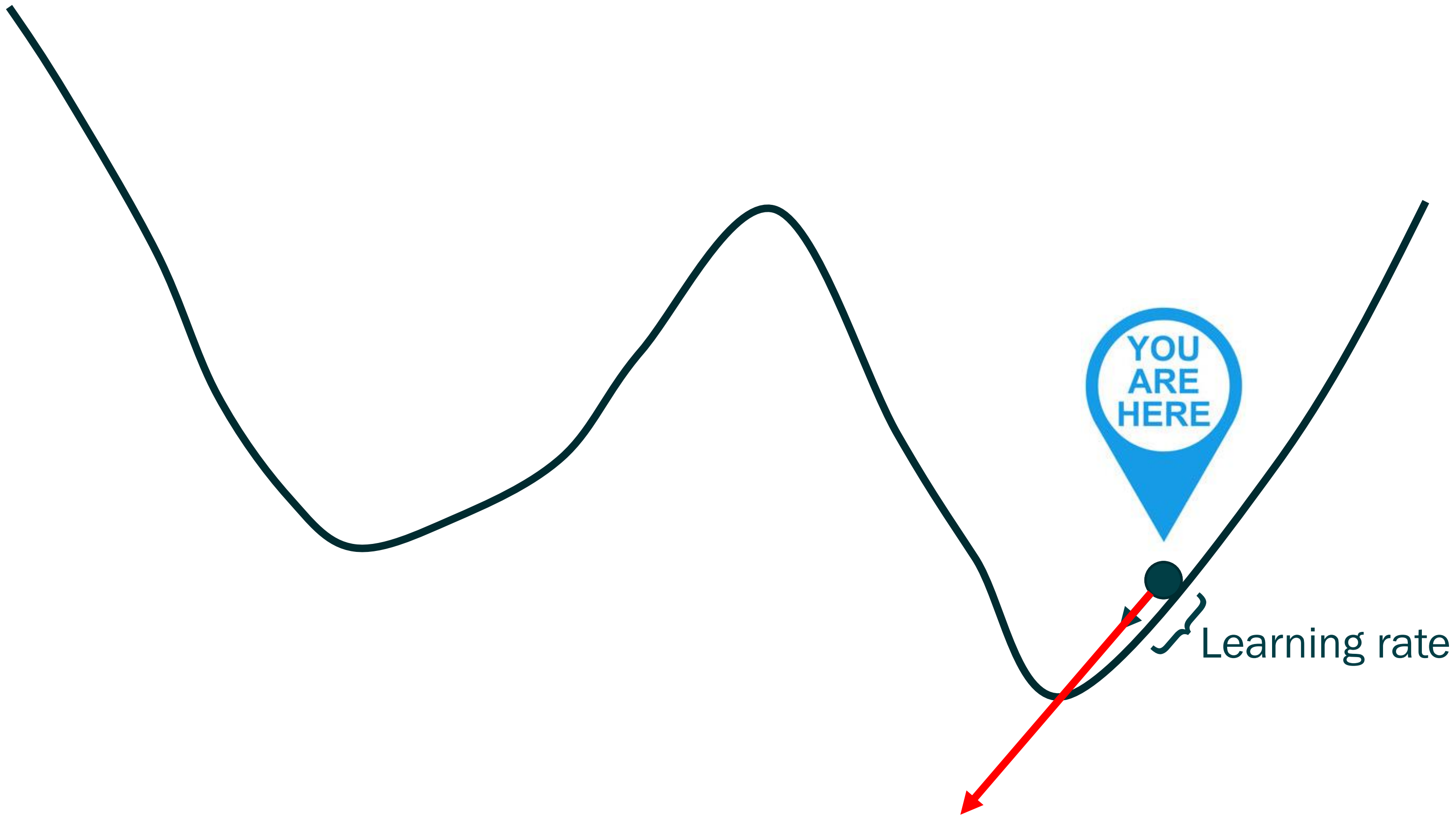
VS.



Can we enable some images to have more influence?

- In today's deep learning, each datum appears only once per epoch during training
- Loss L = $\boxed{\frac{1}{N}\sum_d error(d)} \rightarrow \boxed{\frac{1}{N}\sum_d mass(d) * error(d)}$, where $\sum_d mass(d) = N$
  - Typically, mass(d) = 1 for all d → average error
- What if we increase the mass of some data points vs. others?
  - Mr. Muttonchops gets mass k, where all other data points get mass $\frac{N-k}{N-1}$
- Gradient pushes k times as hard on Mr. M
- Sounds reasonable, right?

Nope. Making some gradients bigger is a bad plan

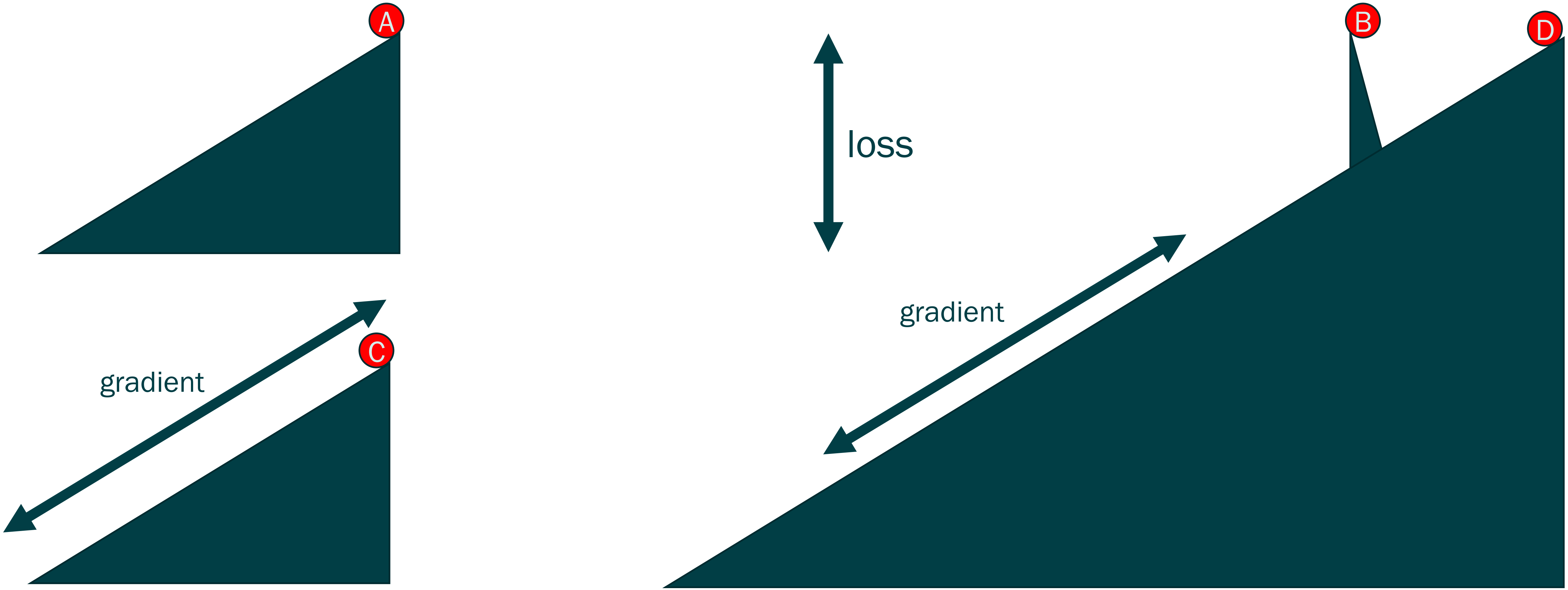


A new idea: repeated selection vs. higher "learning rate"

- If d's relative mass = k , <u>include d (once) in each of ~k minibatches of each epoch</u>
  - Look at d more than once per epoch (in different local contexts)
- Now, d's learning rate is the same as others', but…
- d moves ~k times as far per epoch

- Wait a minute! How should we compute the mass of each datum?
- Loss_d quantifies d's *distance from happiness*: i.e., $\text{loss}_d = 0$
  - Lots of papers advocate Loss_d as relative importance…
- Gradient_d quantifies d's current *velocity on the path to happiness*
  - Lots of papers advocate Gradient_d as relative importance…

Why loss and gradient are poor choices for mass



A new idea: "time to happiness"

- Distance = rate * time → Time = distance / rate
- $T_d = \frac{\|Loss_d\|}{\|Gradient_d\|}$
- Want every data point to achieve happiness at (roughly) the same time
  - Otherwise, either stop before every data point is happy, …
  - Or wait for eons
- Make each datum's mass be equal to its time to happiness
- Datum with more "work to do" gets more time to do it
- Time to happiness is a better criterion than $Loss_d$ or $Gradient_d$

Remassing is powerful

- Optimizes worst-case accuracy, rather than average accuracy
- No customer really cares about average accuracy, yet everybody optimizes that!
  - “Accuracy: Beware of Red Herrings and Black Swans” – Embedded Vision 2020
- But wait! There’s more!
- Remassing can massively accelerate training
  - Focus optimization effort on points with the most work to do
- Most data points resemble other data points: get optimized “for free”!

Facing up to bias

- Remassing optimizes worst-case accuracy, not average accuracy
- Treats rare data points and common data points as <u>equally important</u>
- Treats rare (explanatory) features and common features as <u>equally important</u>

- Remassing addresses a major source of observed bias in face recognition

Resources

Remassing based on gradient direction

https://arxiv.org/pdf/1803.09050.pdf

Remassing based on loss

https://arxiv.org/pdf/1511.06343.pdf

Perceive

https://www.perceive.io

2021 Embedded Vision Summit

"TinyML Is Not Thinking Big Enough" (talk)