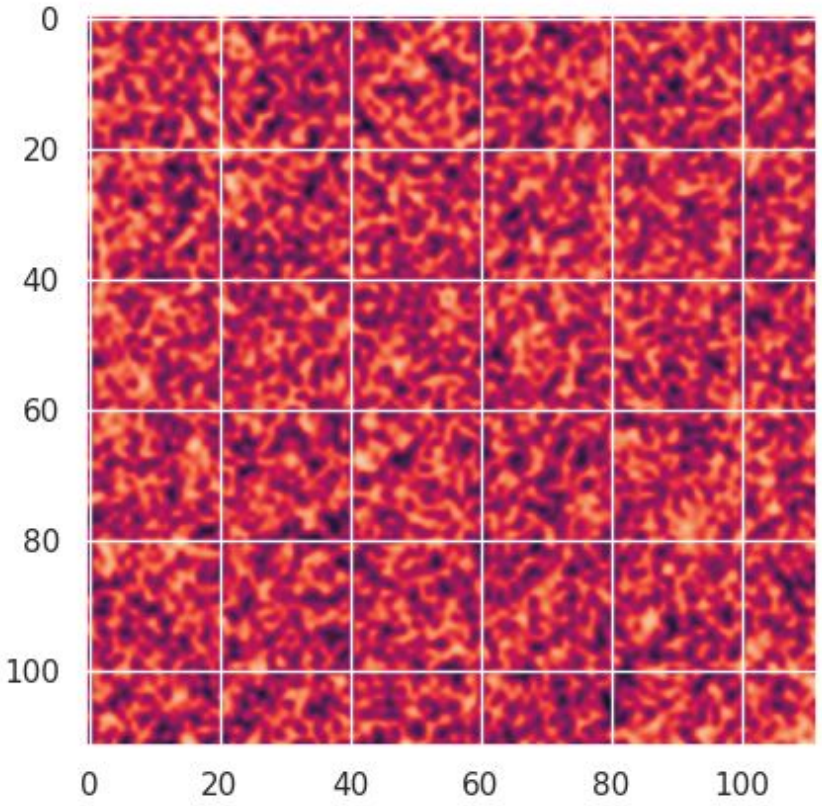
Differential Privacy Applied to Facial Recognition

Source: DALL-E

Kusam Brar, Jared Feldman, Hema Sundaram

December 2023

Who is this?

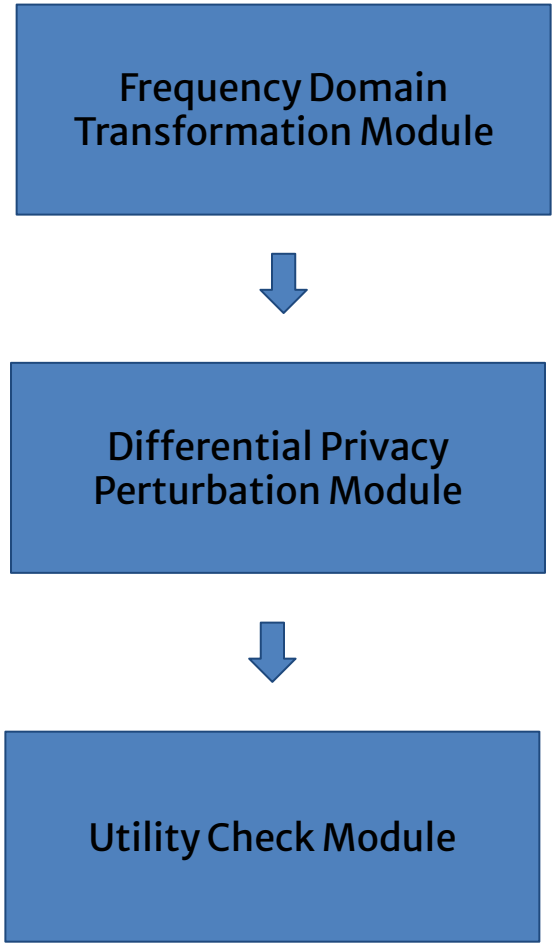


Related Work

- Homomorphic encryption
- Discrete Cosine Transform
- Learnable privacy budgets — our motivation

Datasets

Source: VGG-Face2: A dataset for recognising faces across pose and age.

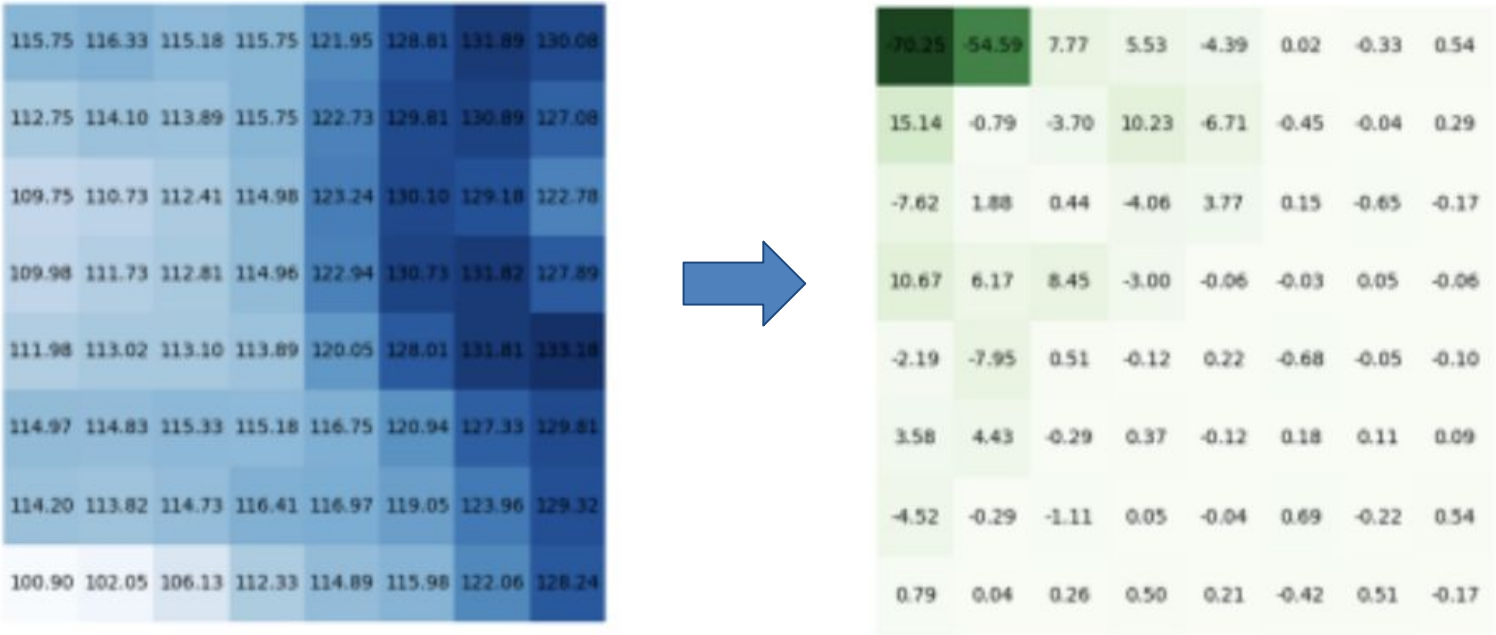

Frequency Domain Transformation Module
Differential Privacy Perturbation Module
Utility Check Module

Frequency Domain Transformation Module

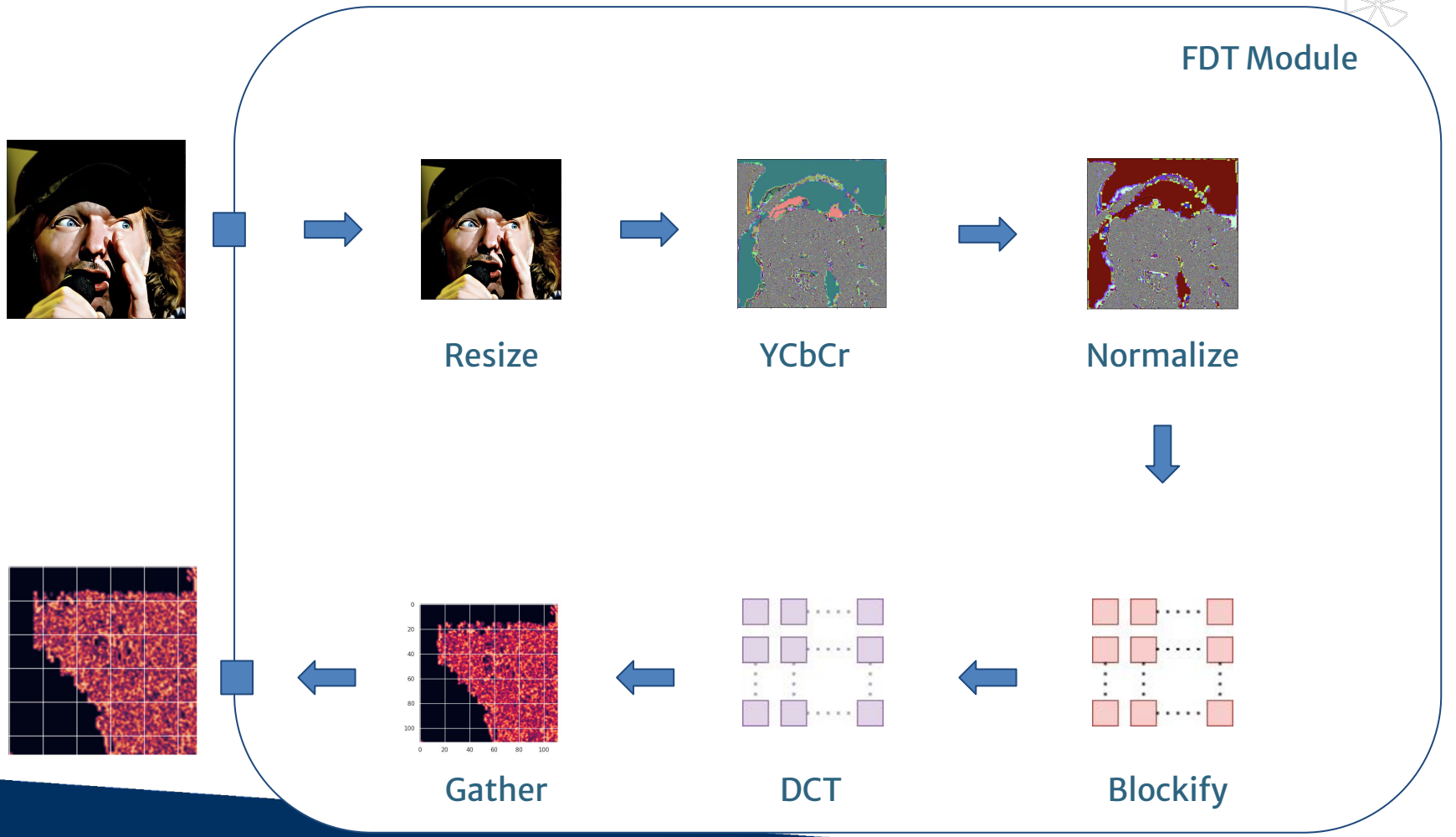
- Preserve visual privacy
- Humans rely on low-frequency information for image recognition
- Neural Networks rely on low and high frequency information for IR
- Using DCT to identify and alter human-readable aspects of images

Discrete Cosine Transform (DCT)

- Transforms signals from one representation to another
- Common use is in compression for JPEG



Frequency Domain Transformation Module



Differential Privacy Perturbation Module

- DCT images were used as input into this model
- Traditional DP vs. image DP
- Element wise distance:

$$d_{i,j,k}(x_1, x_2) = \frac{|x_1 - x_2|}{r_{max}^{i,j,k} - r_{min}^{i,j,k}} \qquad \forall x_1, x_2 \in R_{i,j,k}$$

- Whole representation distance:

$$d(X_1, X_2) = \max_{i,j,k}(d_{i,j,k}(x_1, x_2))$$
$$\forall X_1, X_2 \in \mathbb{R}^{H,W,C}$$

Utility check Module — CNN Model

- Used a model pre-trained on unperturbed images
  - 733 images from 12 different people from different ethnicities
- The perturbed images were validated against the model and accuracy was measured

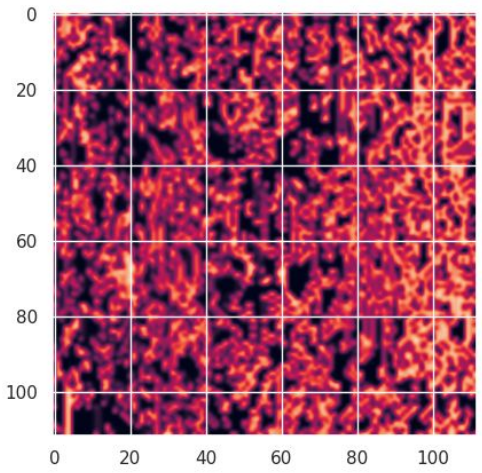
Privacy Preservation and Image Recognition

Results

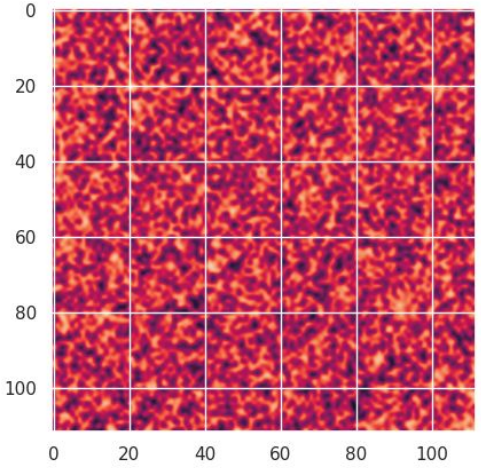
Privacy Preservation

- Image post-DCT unreadable to human eye
- DCT image with noise prevents attacker from reconstructing the original image

Image Post-DCT



DCT Image with Noise



Results — Image Recognition Accuracy

CNN Model Accuracy Comparison	
Baseline Images	Images with Privacy Implementation
72.58%	13.51%

- Privacy/utility trade-off
- Utility better than random guess
- Future Work

Who was that celebrity?

Discussion

Who is it?

Original Image

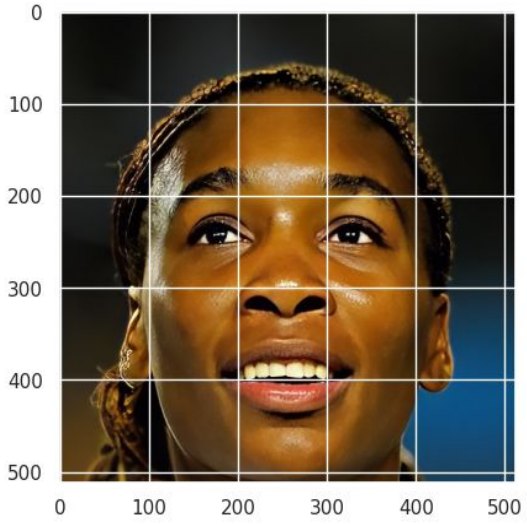
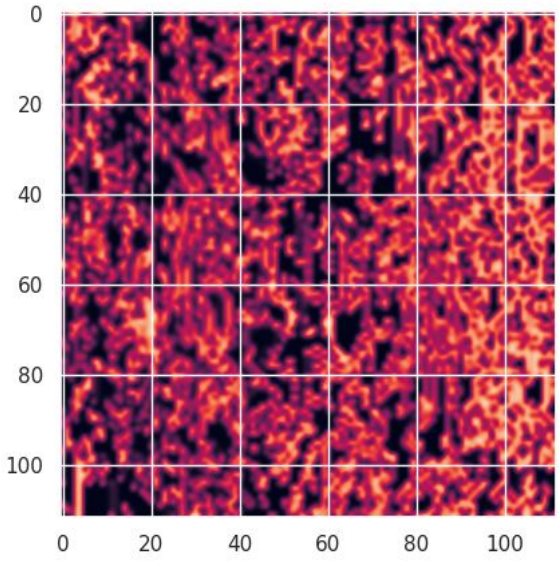
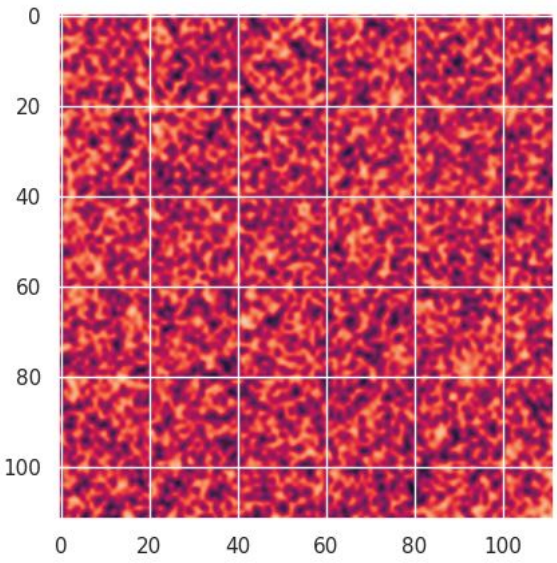


Image Post-DCT



DCT Image with Noise



Venus Williams

Image source: VGG-Face2: A dataset for recognising faces across pose and age.

References

AI TutorMaster. (2023, January 18). *Thought leadership from the most innovative tech companies, all in one place. What is Gaussian Noise in Deep Learning? How and Why it is used?* https://plainenglish.io/blog/what-is-gaussian-noise-in-deep-learning-how-and-why-it-is-used

Alankrita Aggarwal, Mamta Mittal, & Gopi Battineni. (2021). *Generative adversarial network: An overview of theory and applications*. https://www.sciencedirect.com/science/article/pii/S2667096820300045

Elise Devaux. (2022). What is Differential Privacy: Definition, mechanisms, and examples. *Statice.Ai*. https://www.statice.ai/post/what-is-differential-privacy-definition-mechanisms-examples#:~:text=Definition%20of%20differential%20privacy,any%20individual%20in%20the%20dataset

Jahd Khalil. (2023, August 16). Real time crime centers, which started in bigger cities, spread across the U.S. *NPR*. https://www.npr.org/2023/08/16/1194115202/real-time-crime-centers-which-started-in-bigger-cities-spread-across-the-u-s

Jeremy Binckes. (2023, November 15). Yes, people are still using 'password' for their password. *Msn.Com*. https://www.msn.com/en-us/money/other/yes-people-are-still-using-password-for-their-password/ar-AA1jXZSW

Jiazhen Ji, Huan Wang, Yuge Huang, Jiaxiang Wu, Xingkun Xu, Shouhong Ding, ShengChuan Zhang, Liujuan Cao, & Rongrong Ji. (2022). Privacy-Preserving Face Recognition with Learnable Privacy Budgets in Frequency Domain. *ECCV 2022: Computer Vision – ECCV 2022*, 475–491.

Jon Russell. (2017, September 4). Alibaba debuts 'smile to pay' facial recognition payments at KFC in China. *TechCrunch*. https://techcrunch.com/2017/09/03/alibaba-debuts-smile-to-pay/

Joy Buolamwini & Timnit Gebru. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline

Kashmir Hill. (2020, August 3). Wrongfully Accused by an Algorithm. *New York Times*. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

Kay L. Ritchie, Charlotte Cartledge, Bethany Growns, An Yan, Yuqing Wang, Kun Guo, Robin S. S. Kramer, Gary Edmond, Kristy A. Martire, Mehera San Roque, & David White. (2021, October 13). Public attitudes towards the use of automatic facial recognition technology in criminal justice systems around the world. *PLoS ONE*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8513835/

Ken Cabeen & Peter Grant. (n.d.). *Image Compression and the Discrete Cosine Transform*. College of the Redwoods. https://www.math.cuhk.edu.hk/~lmlui/dct.pdf

Liyue Fan. (2018). *Image Pixelization with Differential Privacy*. https://link.springer.com/chapter/10.1007/978-3-319-95729-6_10

M.A.P. Chamikara, P. Bertok, I. Khalil, D. Liu, & S. Camtepe. (2020). Privacy Preserving Face Recognition Utilizing Differential Privacy. *Computers & Security*, *97*. https://www.sciencedirect.com/science/article/pii/S0167404820302273

Mei Wang & Weihong Deng. (n.d.). *Ethnicity Aware Training Datasets* [dataset]. http://www.whdeng.cn/RFW/Trainingdataste.html

Nick Galov. (2023, May 20). 20 Facial Recognition Statistics to Scan Through in 2023. *Web Tribunal*. https://webtribunal.net/blog/facial-recognition-statistics/

Oloid Desk. (2023, November 6). Facial Authentication Revolution: 15 Industries Embracing the Future of Security. *OLOID*. https://www.oloid.ai/blog/facial-authentication-revolution/

Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, & Andrew Zisserman. (2018). *VGGFace2: A dataset for recognising faces across pose and age*. https://www.robots.ox.ac.uk/~vgg/publications/2018/Cao18/cao18.pdf