Don't Take Things at Face Value: Facial Age Classification

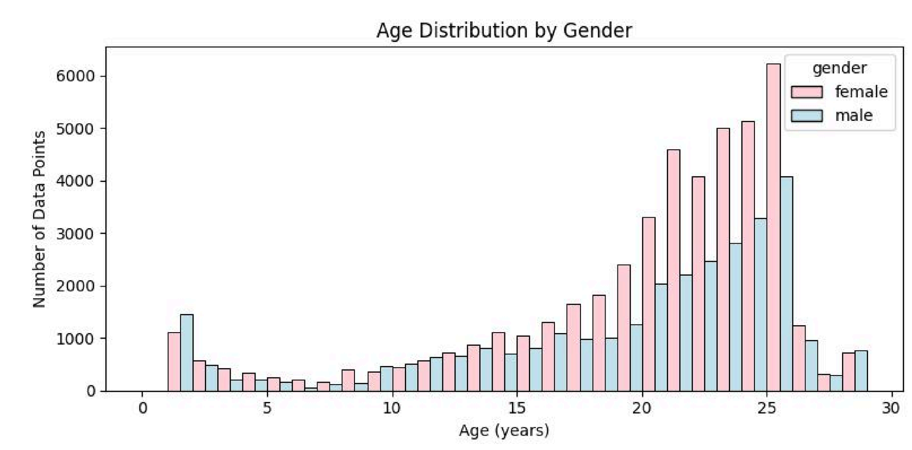
Deeksha Vatwani, Sriya Bulusu, Lam Pham

Motivation

- **Enforcing age restrictions** is vital for both user safety and regulatory compliance on digital platforms
- **Current methods** like self-reported birth-dates and ID uploads are **easily bypassed**
  - 45% of under-13s access social media
  - ~50% of underage alcohol orders succeed
- **Human annotators struggle with age estimation** due to subjective biases (e.g., cultural differences in aging signs)
  - CNNs can objectively identify discriminative features (e.g., wrinkles, facial proportions) across age groups
- Existing deep learning approaches to age classification do not target **critical verification boundaries (e.g., 13, 18, 21)**
- Can a CNN trained to perform facial age estimation be a non-intrusive, scalable approach to enhance both accuracy and user experience in age verification?

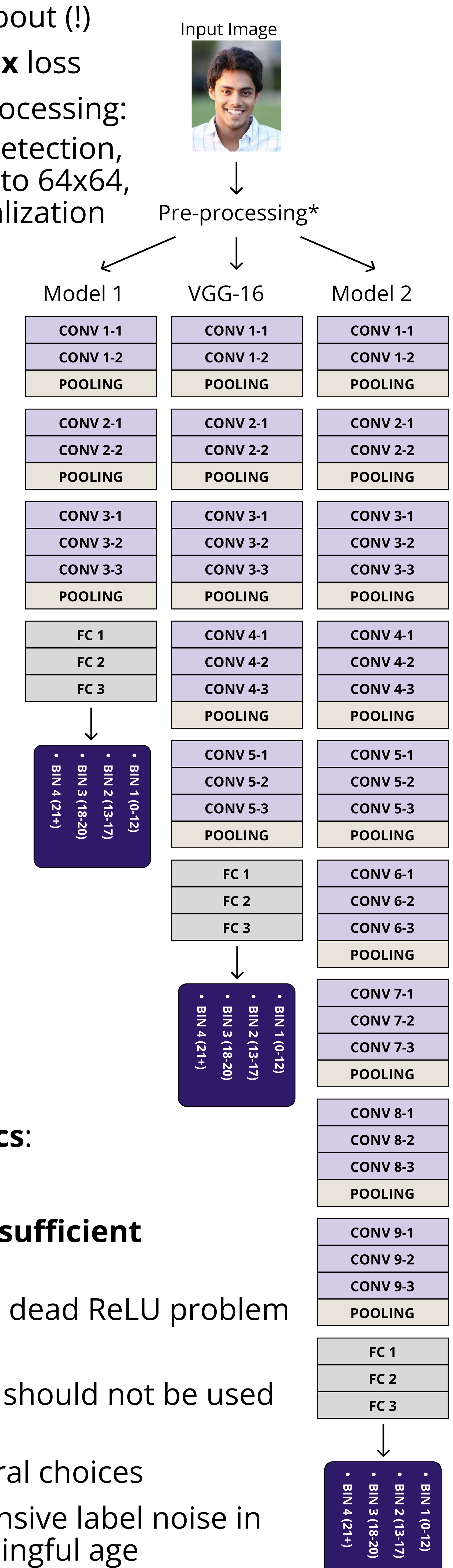
Datasets

- There are large, well-known, publicly available datasets of faces with ages labeled (e.g., AAF, Adience, AFAD, MORPH); we use a mix of:
  1. **UTK-Face:** 23,708 images, ages 0-116, diverse ethnicities and poses
  2. **IMDB-WIKI:** 500,000+ celebrity face images with estimated ages from metadata
- **Limitations:**
  - Under-representation within 9-25 age range, noisy mislabeling, distribution bias towards older adults (20-50 years), celebrity photo scenarios
- Created **70/20/10 train/val/test split** on combined, balanced dataset
- Very difficult to procure 9-25 age data: legal/ ethical barriers, annotation difficulty



Setup & Architecture

- **Three CNN model architectures**, inspired by VGG-16, but with varying depths
  - **ReLU** after each CONV layer
  - **Optimizer:** Adam (lr=1e-4)
  - **Batch size:** 32
  - **Training epochs:** 10
  - **3x3** kernel sizes, **Maxpool** layers
  - No dropout (!)
  - **Softmax** loss
  - *Pre-processing: Face detection, resize to 64x64, normalization



Results

	Per-Bin Accuracy				
	< 13	13 - 17	18 - 20	21+	Overall
Model 1 - Simple	23.51%	50.15%	26.71%	35.62%	35.54%
VGG-16	29.73%	43.05%	24.46%	41.31%	37.80%
Model 2 - Deep	0.00%	100.0%	0.00%	0.00%	15.66%

Table 1. Test accuracy of each model, segmented by age bins.

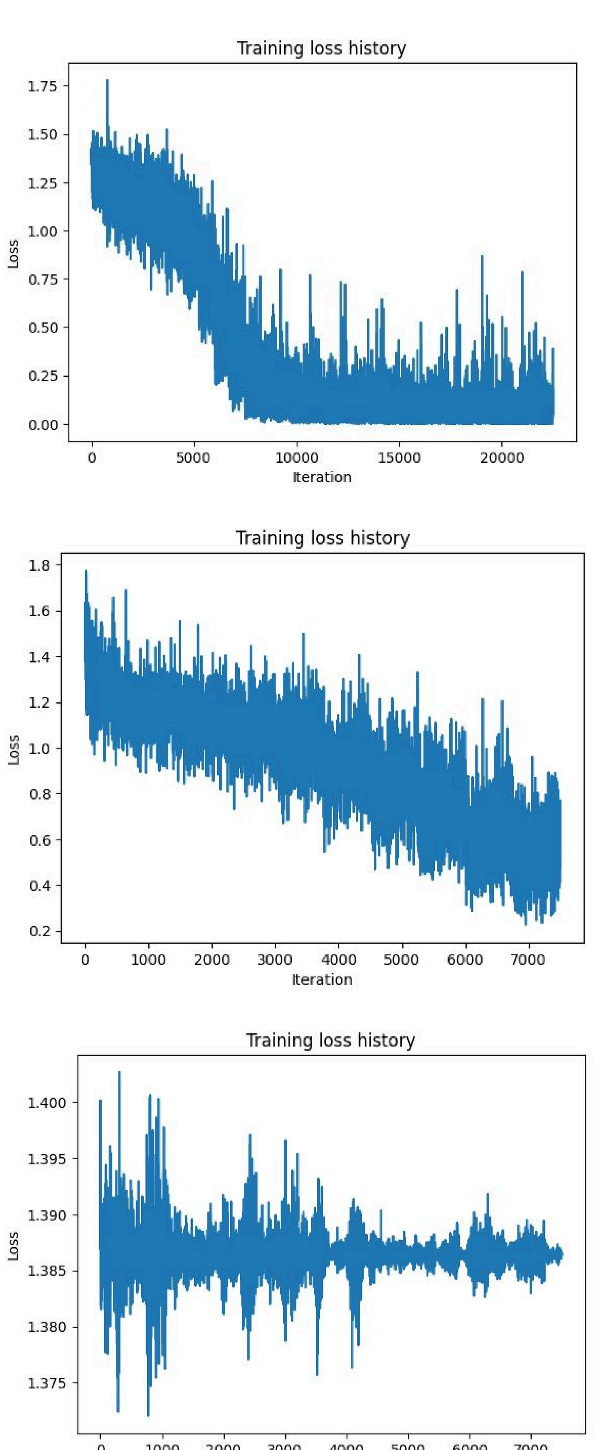
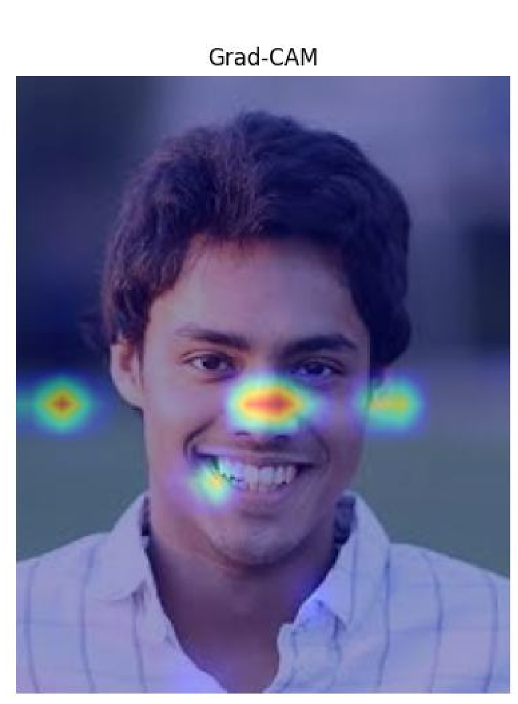


Figure 1. Training loss history of each model, ordered: Model 1, VGG-16, Model 2



Model 1: 13-17

VGG-16: 21+

Model 2: 13-17

Discussion

- All three models achieved **poor test accuracy** (<40% overall), with performance varying dramatically across age bins
  - Contrasts with very high training accuracy (90%+), indicating severe **overfitting**
- Grad-CAM analysis exposes that **models are NOT learning facial age features**:
  - Model 1: Random, scattered activation patterns suggest **spurious correlation learning**
  - VGG-16: Focus on **clothing/background** rather than facial characteristics
  - Model 2: Uniform activation maps indicate gradient saturation and **learning failure**
- Loss curves reveal **concerning patterns in training dynamics**:
  - Model 1: Rapid **overfitting** with erratic validation loss
  - VGG-16: Stable but high loss, indicating **insufficient adaptation** to age features
  - Model 2: **Collapsed training**, likely due to dead ReLU problem within deeper architecture
- (Our) facial age classifier models *definitely* should not be used as unilateral decision-makers, BUT…
  - **Dataset limitations** trump architectural choices
  - Severe under-representation and extensive label noise in existing standard datasets make meaningful age discrimination near **impossible within 9-25 age range**
- Resort to **memorizing** minimally available images OR learning **non-facial cues** (clothing, background), but fail to generalize

Future Work!

- Curate a larger, more application-accurate dataset *without* mislabels
- Perform more applications/analysis (e.g., Botox, makeup, lighting)
- Explore attention mechanisms to focus on age-critical regions

References

[1] Danah Boyd, Eszter Hargittai, Jason Schultz, and John Palfrey. Why parents help their children lie to facebook about age: Unintended consequences of the children's online privacy protection act. First Monday, 16(11), Oct. 2011.
[2] Gil Levi and Tal Hassncer. Age and gender classification using convolutional neural networks. In 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 34–42, 2015.
[3] Julia Martinez, Patricia Rutledge, and Kenneth Sher. Fake id ownership and heavy drinking in underage college students: Prospective findings. Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors, 21:226–32, 06 2007.
[4] Muhammad Mustapha, Nur Mohamad, Ghazali Osman, and Siti Ab Hamid. Age group classification using convolutional neural network (cnn). Journal of Physics: Conference Series, 2084:012028, 11 2021.
[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[6] Rebecca S. Williams and Kurt M. Ribisl. Internet alcohol sales to minors. Archives of Pediatrics Adolescent Medicine, 166(9):808–813, 09 2012.
[7] Song Yang Zhang, Zhifei and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.
[8] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision, 126(2-4):144–157, 2018.