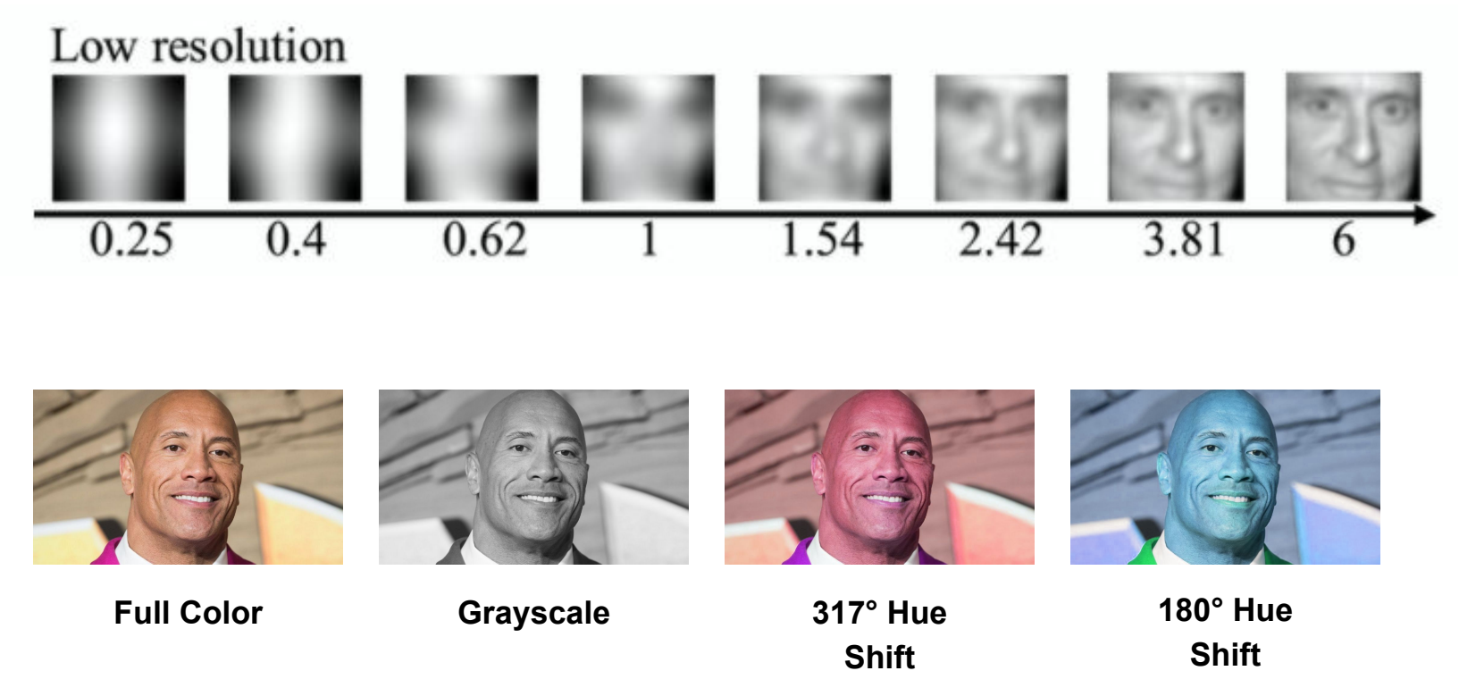
Transformation Tolerance of Machine-based Face Recognition Systems

By: Ashika Verma[1,2,3], Kyle Keane[1,2], Alyssa Unell[1], Anna Musser[1], Pawan Sinha[1]

[1]MIT Brain and Cognitive Sciences Department, Sinha Lab
[2]MIT Schwarzman College of Computing, Quest for Intelligence
[3]MIT Dept of Computer Science and Engineering

Introduction:

- Facial Recognition is one of the most widely used biometric authentication methods in the world today
- Humans have a remarkable ability to accurately recognize faces under a variety of naturalistic degradations, such as grayscale, pseudocolor, and blurred images
- In humans, color only plays a significant role in recognition under blurred conditions (Yip) (Fig.1)



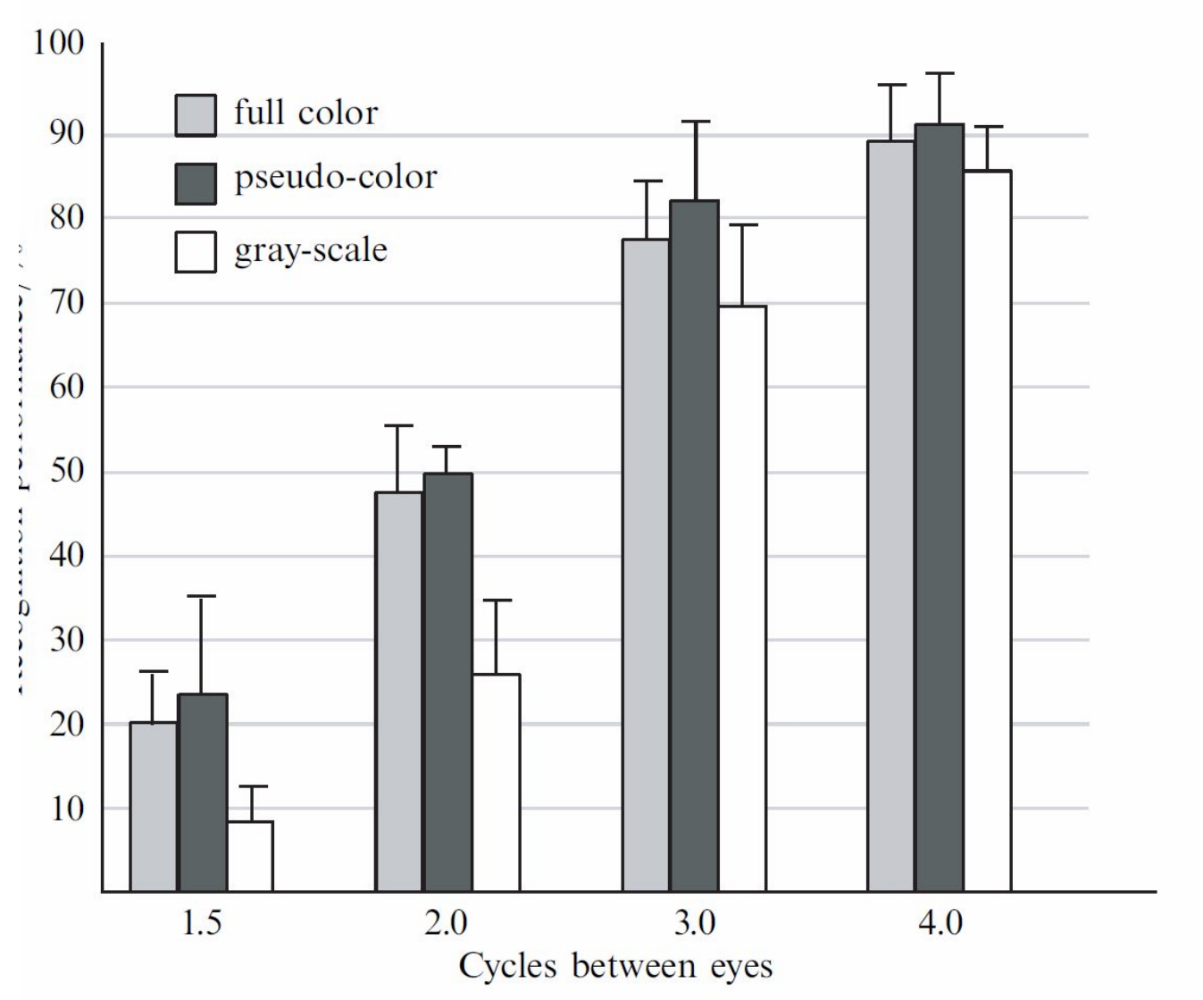


Figure 1: Results of human study regarding biological ability to recognize degraded faces

- The impact of these degradations on facial recognition for state-of-the-art networks has not yet been explored
- We gained a deeper understanding of facial recognition in machine learning and neuroscience by comparing the impact of degradations

Methods:

- Used ResNet-101 trained on Augmented CASIA-WebFace Data
- Tested on CelebAMask-HQ dataset variations that underwent degradation compositions
  - Example gray, blur, pseudocolor
- The images were passed into network and encoded into a vector space (Fig. 2)

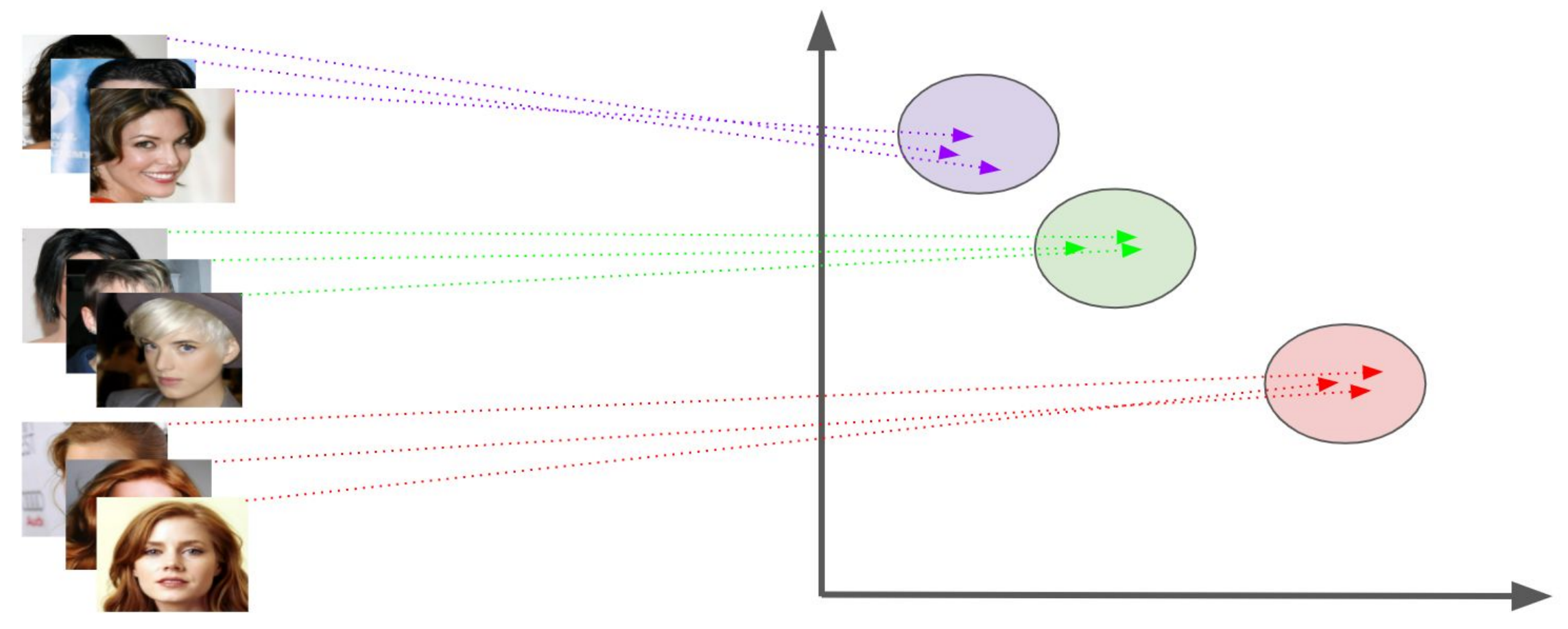
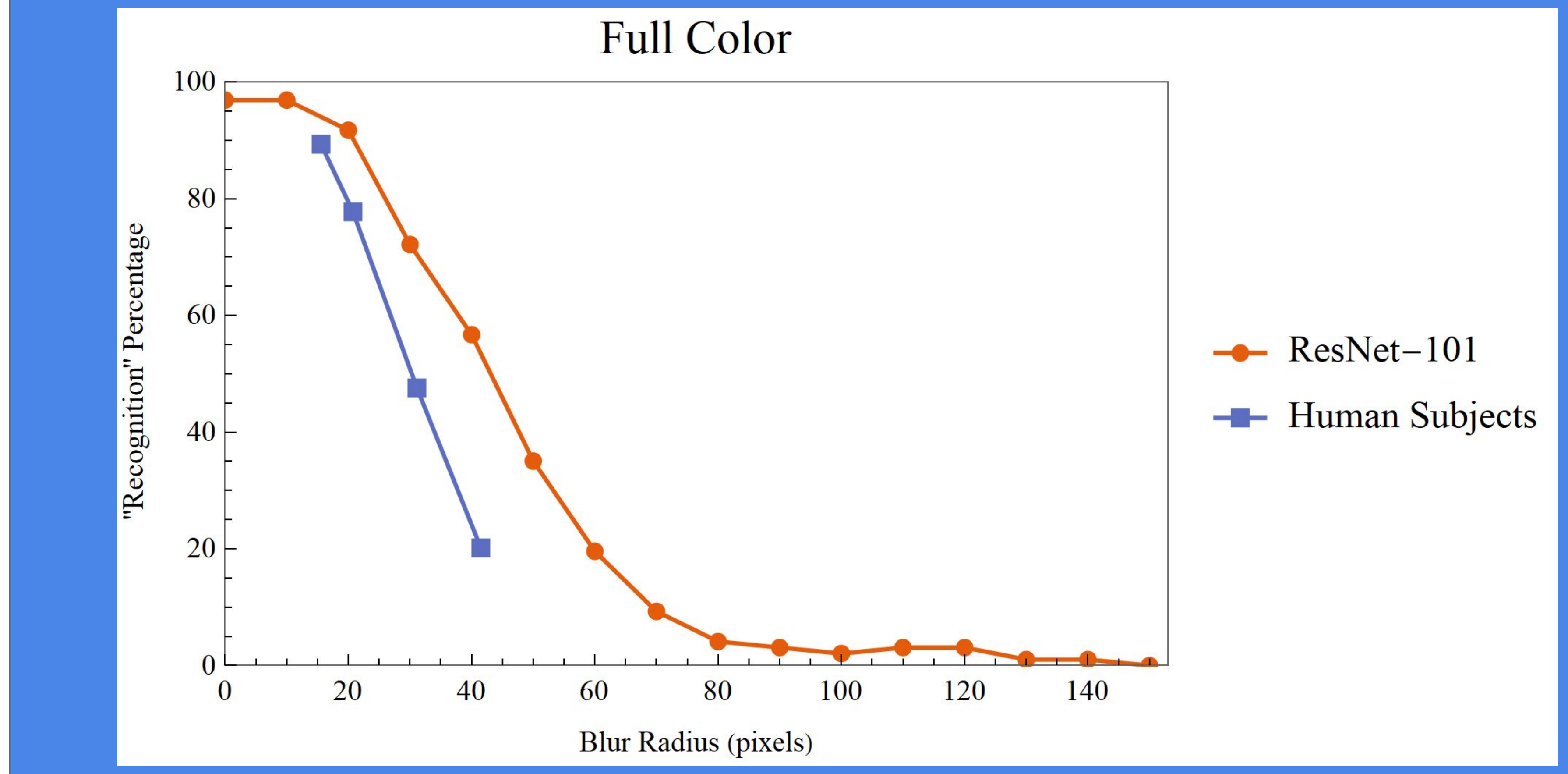


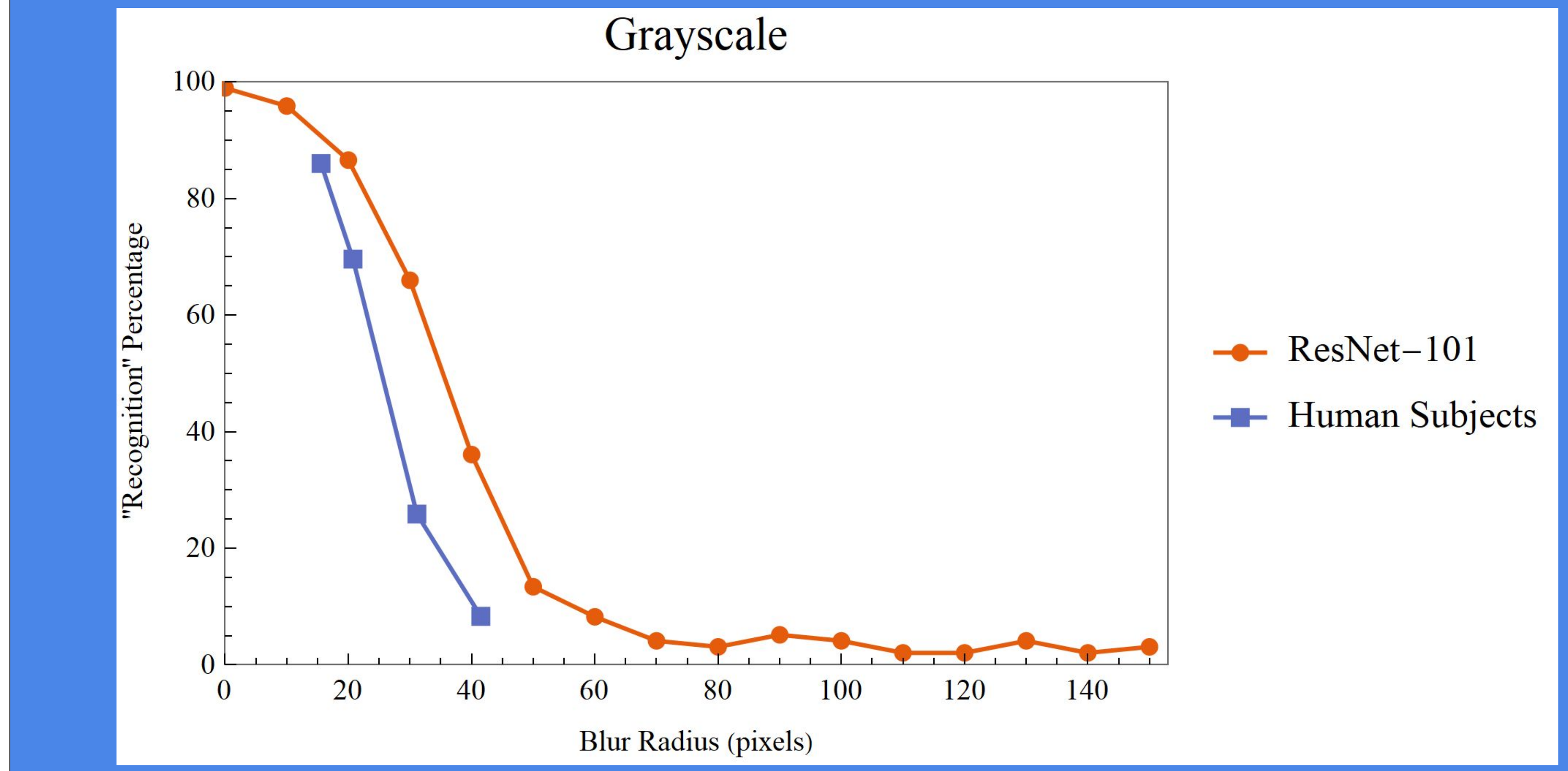
Figure 2: Example visualisation of image encoding in vector space

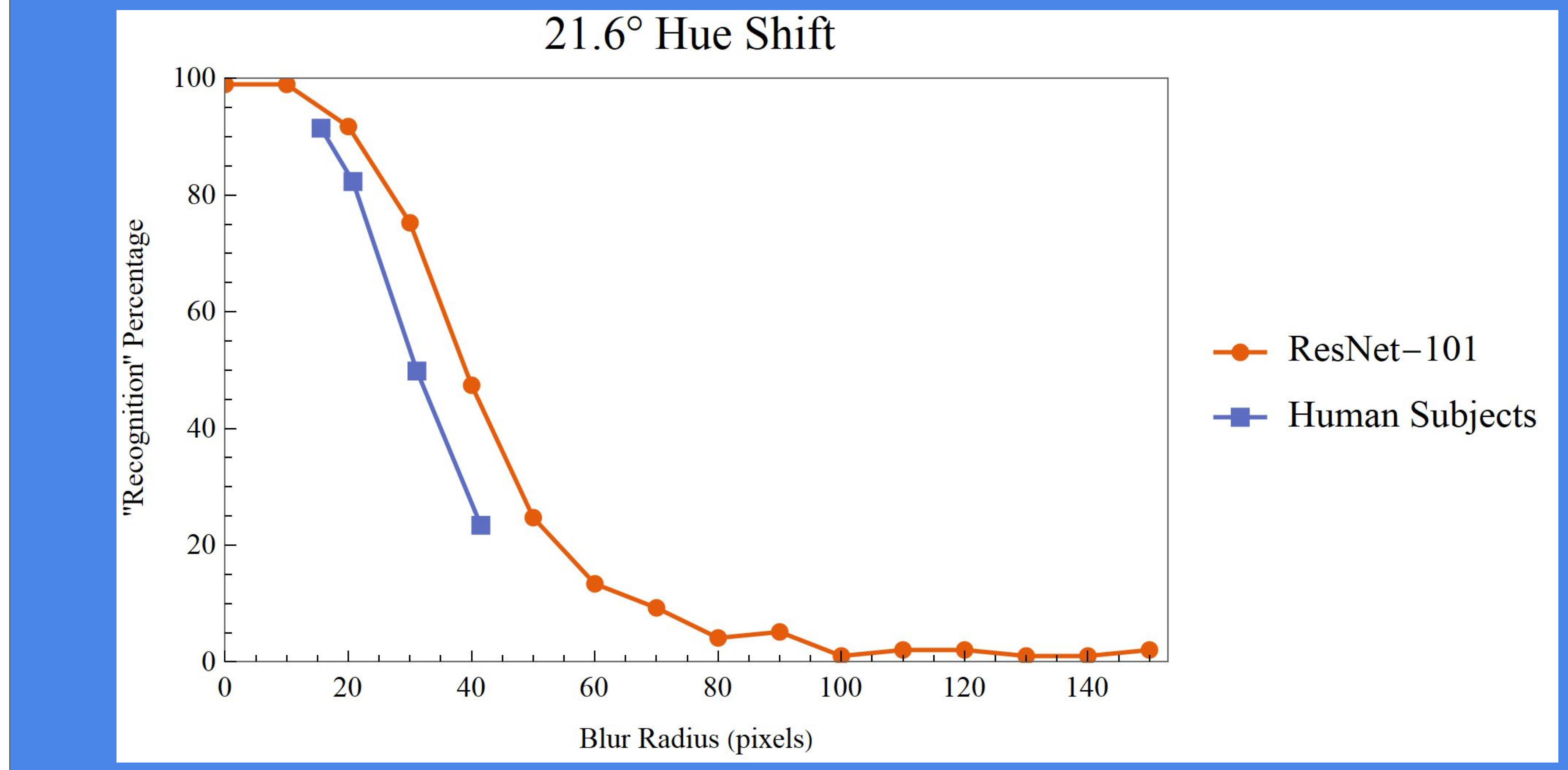
- Images are classified correctly when they are, on average, closer in vector space to all other images that are the same identity according to ground truth (Fig. 3)
- Overall an identity is “recognized” if 75% of that identity’s images are, on average, closer to images of that identity than to images of any other identity.

Color plays a **significant role** in the accuracy of facial recognition in AI systems under **blurred conditions**, matching human studies

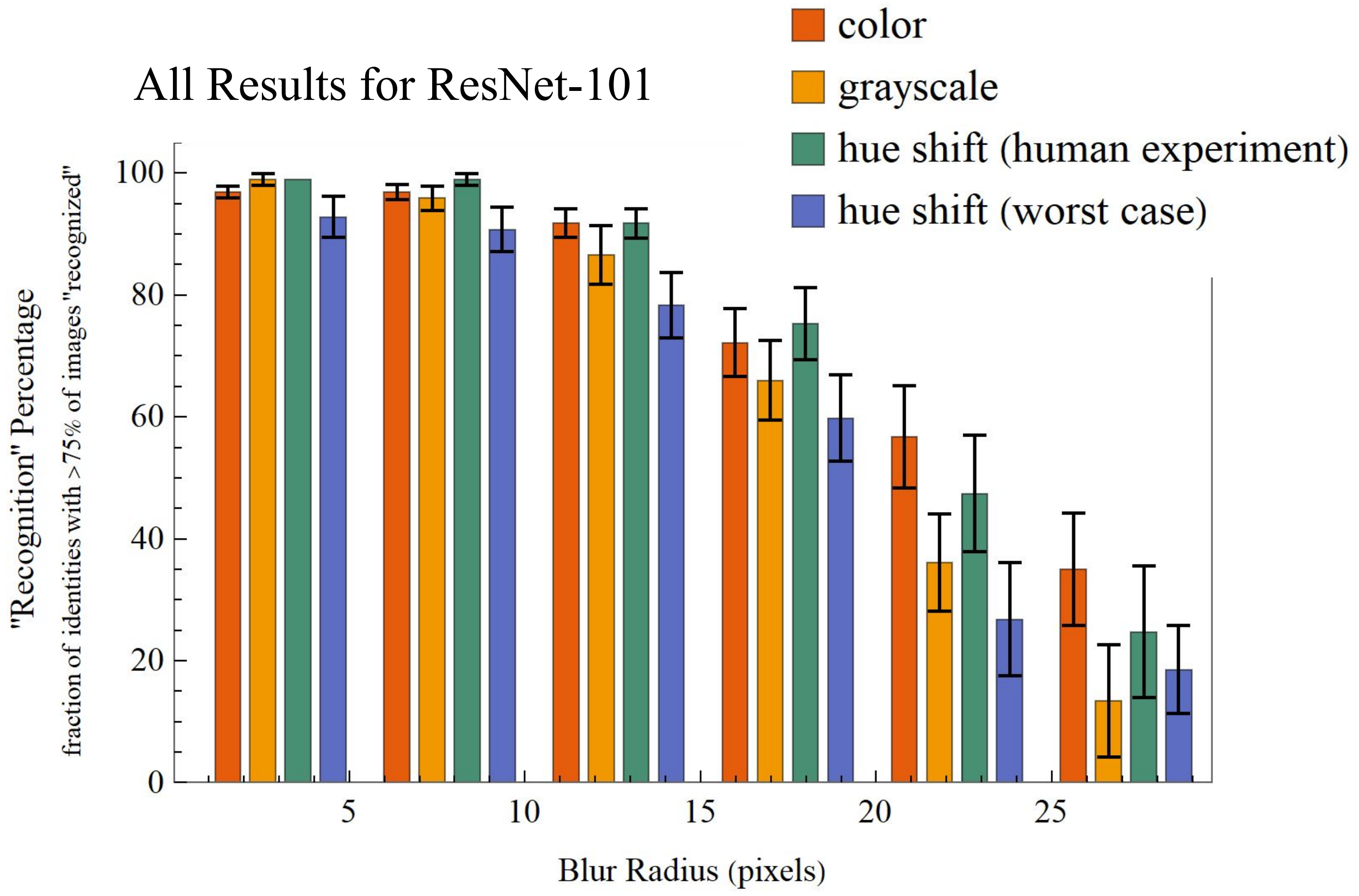
Human Results Compared to ResNet-101





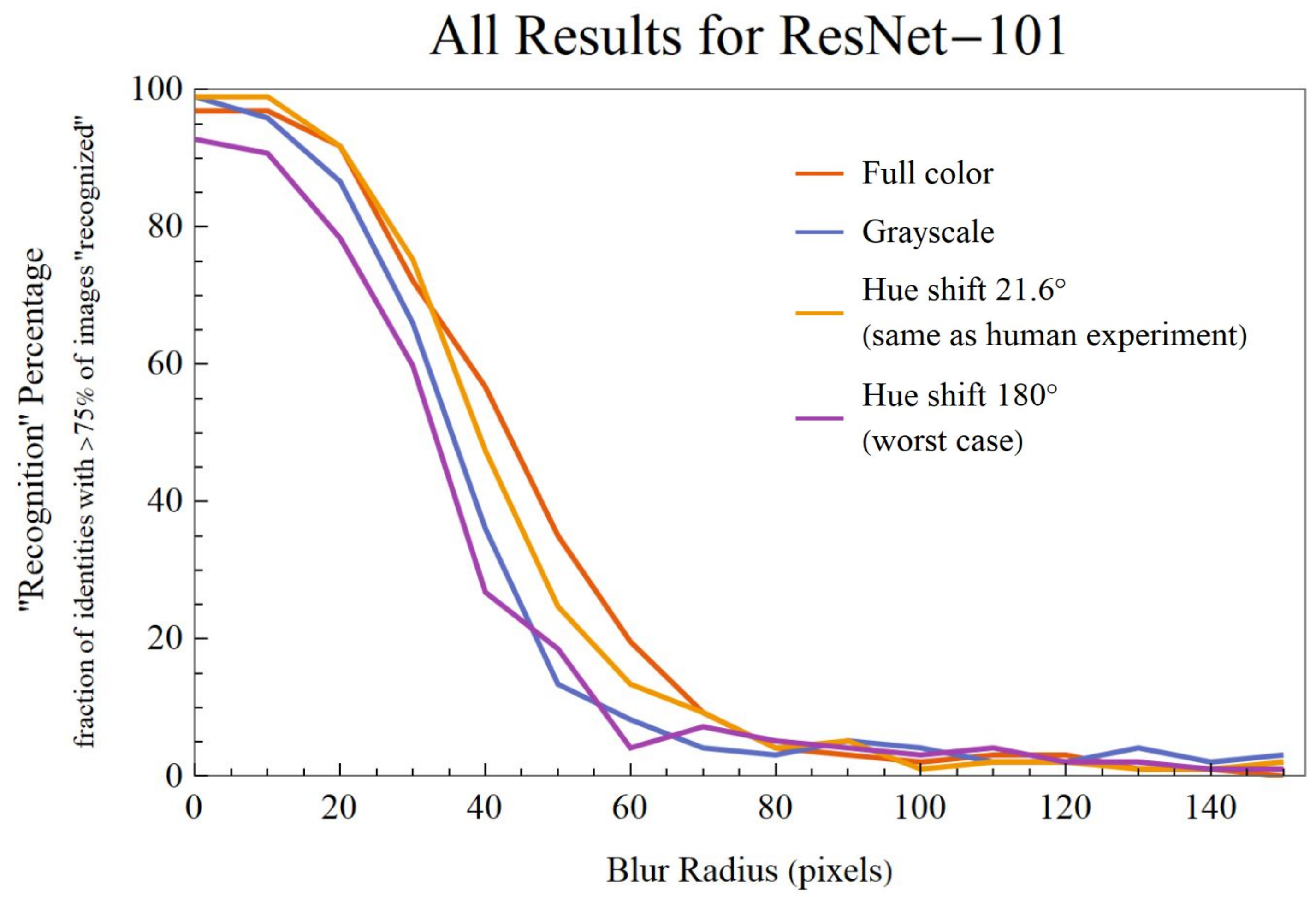


Results:



Key Results:

- The network performs the best on full color, then on a 21.6° hue shift, then the worst case hue shift (180°), then grayscale
- Color alone doesn’t contribute to better facial recognition



Discussion:

- Why is our network working like this?
  - The neural network was trained on full color images, we expected to see the neural network to do worse on any kind of hue shift and grayscale degradation
  - For humans, we hypothesize that color is needed to segment out different parts of the face
  - For neural networks, we cannot necessarily say the same; the worst color hue shift dramatically decreases the recognition performance, meaning the original color of the faces are important in its vector encoding
- Some limitations:
  - The diversity of this dataset is limited: mostly Caucasian faces
  - Further work could include neural networks FaceNet and CLIP
  - The human experiment used for comparison was done in 2002 with limited data; it would be ideal to collect more human data using the same data set we used here for the AI system
- Future research:
  - Further comparisons between human and AI performance could be done using different degradations, ex: line drawing, caricatures, photonegative images