Predicting 2D Human Orientation from Images

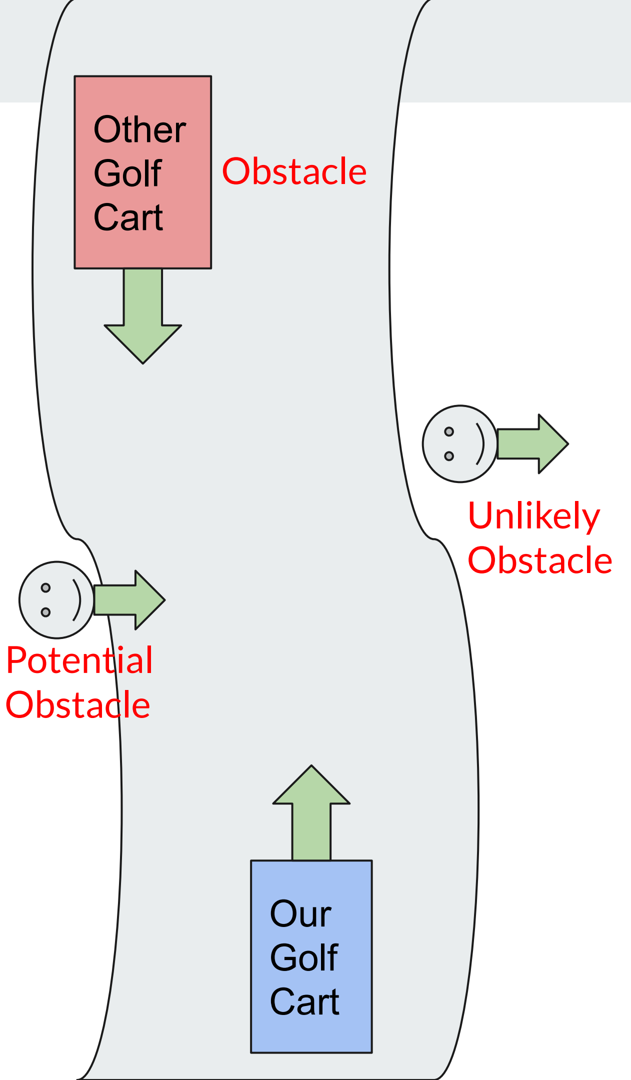
EECS 731: Intro to Data Science

Instructor: Dr. Hongyang Sun

Group 9
Caden Kroonenberg, Joe Nordling, Ahmni Pang-Johnson, Aydan Smith, Nick Velicer, Derek Zhang

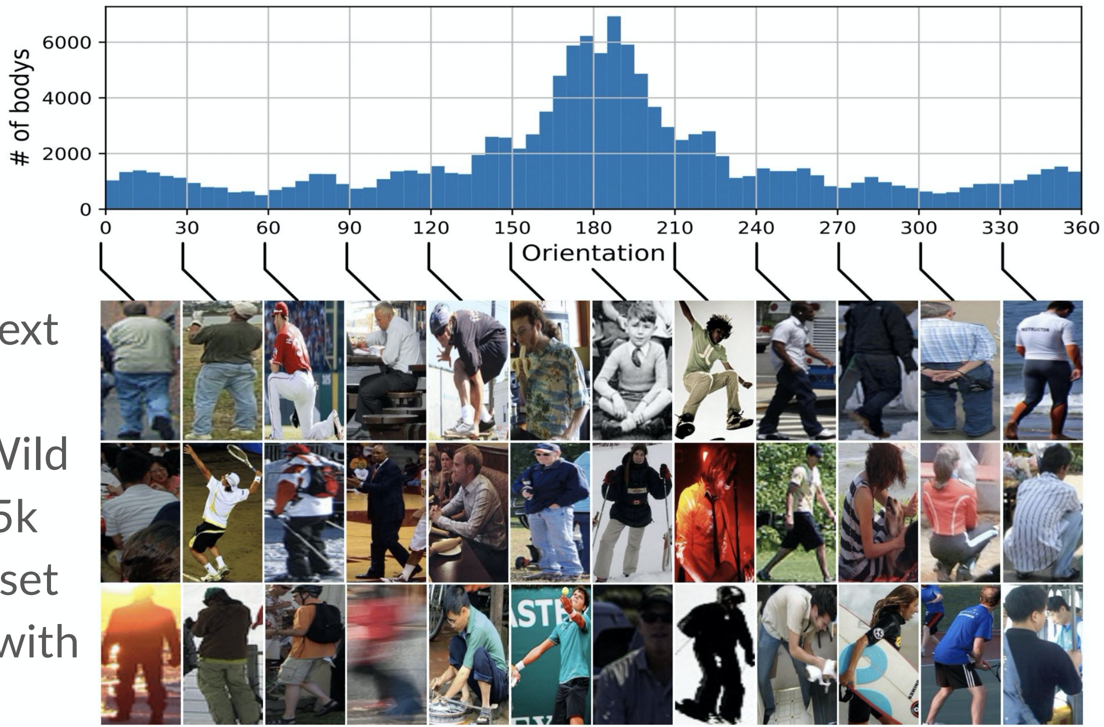
Background

- Self-Driving Golf Cart
- Collision Space:
  1. Detect Objects (People, Vehicles, Etc.)
  2. Classify Orientation
- Use object orientations to predict movement
  - An object is likely to move in the direction it is facing
- Based on predictions, we can generate a safe path in our path planning module



Dataset

- COCO-MEBOW
- Common Objects in Context
- Monocular Estimation of Body Orientation in the Wild
- 130k annotations from 55k images in the COCO dataset
- 72 bins to partition 360° with each bin covering 5°

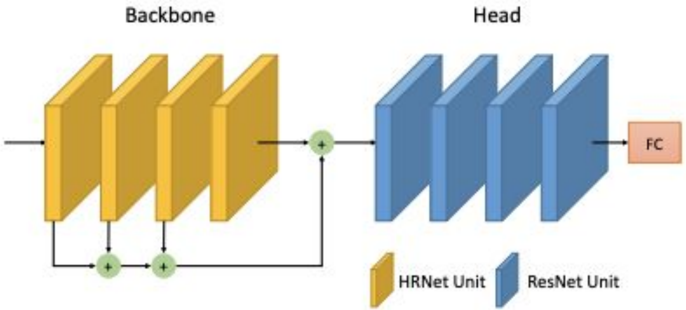


Approach

- Provided Model Provided with Dataset
- ResNet Implementation
- Data Pre-Processing

Provided Model

- Architecture
  - HRNet used as backbone unit
  - ResNet used as head unit
  - 72 bins for orientation
- Implementation
  - HRNet (High-Resolution Net)
    - General Purpose Convolutional Neural Network
    - Able to use high resolution representations of images through the whole process
    - Commonly used for detecting objects and classifying images
  - ResNet (Residual Networks)



Provided Model Performance

	Method	MAE	Acc.-22.5°	Acc.-45°
	AKRF-VW [18]	34.7	68.6	78
	DCNN [19]	26.6	70.6	86.1
	CPOEHK [53]	15.3	75.7	96.8
Provided Model ==>	**ours**	8.4	95.1	99.7
	Human [18]	0.93	90.7	99.3

ResNet Implementation

- Basic structure: Residual Blocks
    - Employ skip-connections to pass information deeper into the network
    - Skip-connections mitigate loss of gradient descent error common in NNs
- Residual Blocks are stacked to create a Residual Network
- Increased performance as network depth increases

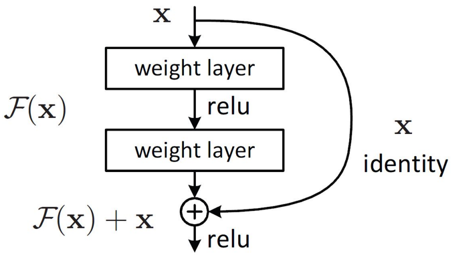
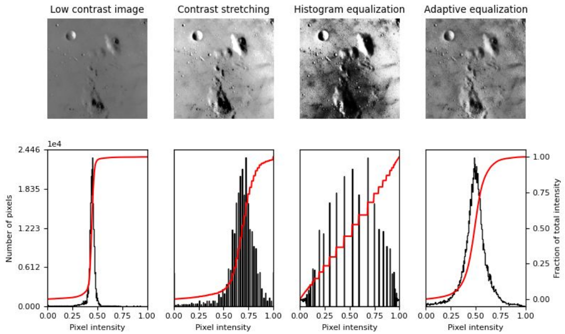


Fig. 1: Residual Block

$$f(x) = x^{+} = \max(0, x),$$

Fig. 2: ReLU function

Data Pre-Processing



1. Optimizing contrast and efficiency by using histogram equalization
    a. Histogram is a frequency distribution that plots the brightness of each pixel, goal is to flatten (equalize) the frequencies of each level of luminance
    b. Pros
        i. Training runtime decreased due to decrease in image information content
    c. Cons
        i. Could make it harder to detect humans due to discoloration and loss of image info
        ii. Can alter brightness of image dramatically

Data Pre-Processing

2. Reduce background noise by applying gaussian blurring
    a. Apply a gaussian kernel to the given image, with the purpose to reduce gaussian noise
    b. Pros
        i. Background noise may be reduced, increasing accuracy in determining human orientation

    Cons
        i. Too much blurring can be counter-intuitive, leading to increased difficulty determining orientation

Data Pre-Processing

3. Use unsharp masking (inverse Gaussian blur) to de-noise and increase feature contrast in the image
With a kernel (see below), apply it to an image, and subtract the result from the original to sharpen at the cost of continuity.

Pros:

- Increases boundary sharpening
- Adds definition to features

Cons:

- Ignores more detailed image information
- Sacrifices continuity for extra contrast

Goal

Compare model performance and effects of pre-processing to have an informed decision of the best approach to implement on self-driving golf cart

Model	Provided Model	ResNet
No Pre-Processing Performance	???	???
Histogram Equalization Performance	???	???
Gaussian Blur Performance	???	???
Unsharp Masking Performance	???	???

References

1. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
2. Wu, Chenyan, et al. "MEBOW: Monocular Estimation of Body Orientation in the Wild." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.