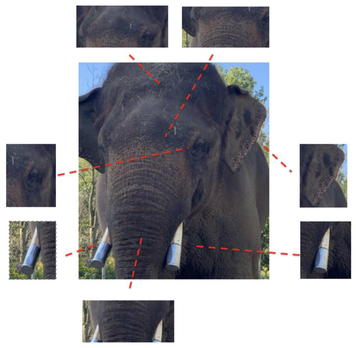
Individual Elephant Face Classification

Shashank Kumar, Ujwala Musku, Rupesh Deshmukh, Luddy School of Informatics and Computing, Indiana University Bloomington

1. Motivation

- The problem statement is to **classify individual elephants based on their facial features** the zoo.
- Applications:
  Cataloging **patterns** of **elephant movements** in the zoo and **enhancing virtual interaction** for visitors by providing background stories and family/species history.



A model to learn facial features such as eyes, trunk, tusk, marks on the ear and trunk.

- We fed annotated elephant face images to multiple convolutional based networks to model elephants based on discriminative facial features.
- We can identify differentiative facial features to provide the **evidence and rationale** for correctly classified and misclassified elephants respectively using saliency maps.

2. Challenges

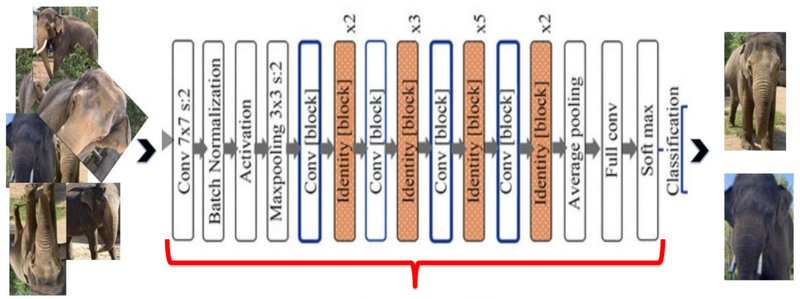
- **Less data per elephant class:** Some elephant classes have as low as 60 images.
- **Variability among images:** Images taken within a short span of time have low variability among them.

- **Common background:** Blue sky or trees appearing in images of multiple classes led to misclassification.

***On the left**: Elephants of two different classes misclassified as being the same due to the common background interpreted as a class feature.*

3. Approach

- The **faces of 18 elephants with 1778 images are extracted** based on the annotations provided.
- For each elephant, the images are clustered into train and test datasets using **Agglomerative clustering** technique based on **pairwise image-hash difference**.
- The images are **augmented** by flipping horizontally, adding sheer and zoom range, height and width shift, and 20 degrees rotation range.
- We used transfer learning with **ResNet50, VGG16, VGG19**, and **Efficient-Net** along with 2 additional trainable fully connected layers and implemented **discriminative learning** in different layers.
- **Accuracy** is measured on cropped and uncropped images for **11** and **18 elephant labels** separately.

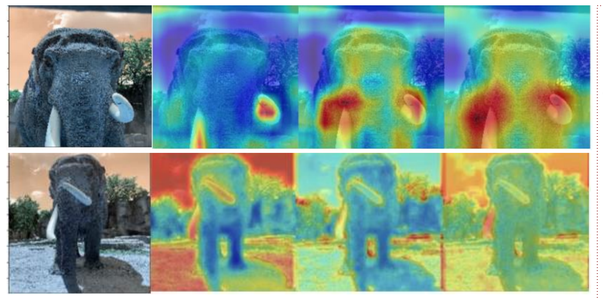


ResNet50

4. Experiments

- Baseline model with 5 convolutional layers gave us very high accuracy of 97% due to high similarity in images. After Agglomerative clustering on data set, the accuracy dropped to around 27%.
- Horizontal flip with adding sheer and zoom range augmentation gave us best accuracy in all transfer learning on ResNet50, VGG16, VGG19,and Efficient-Net for 11 elephant classes.
- The best model(ResNet-50) is picked, and all 18 elephants labels are predicted.

5. Results



TopLeft1: Correct
TopLeft2: Misclassified
Saliency Map
Output of last three convolution layers
In the first image, Tusk was identified as the main feature to classify the image
Second image, background trees and blue skies are creating misclassification

Aug-1 : Default parameters; Aug-2 : Horizontal flip
Aug-3 : HFlip, Rotation range (20 degrees), Height and width shift ; Aug-4 : Sheer, Zoom, and HFlip

	Test Set	No Aug	Augmentation				
			Aug. 1	Aug. 2	Aug. 3	Aug. 4	Aug. 4 w/ Discrim. Learning
Baseline	Cropped	38.650	35.890	68.098	3.988	50.307	-
	Uncropped	26.993	26.380	41.104	3.681	28.834	-
VGG-16	Cropped	82.515	90.491	87.730	43.865	87.117	88.583
	Uncropped	70.859	74.233	75.153	42.025	76.687	78.221
VGG-19	Cropped	87.491	91.104	90.184	40.491	90.491	92.630
	Uncropped	73.129	74.847	76.380	41.104	75.767	80.061
ResNet-50	Cropped	94.479	93.865	92.638	7.055	94.172	95.399
	Uncropped	83.196	84.663	84.356	9.509	86.503	88.344
EfficientNet	Cropped	69.325	66.564	73.313	6.748	96.319	85.890
	Uncropped	60.736	52.761	68.712	8.589	78.528	75.233

***Above:** Classification Accuracies of all tested models over cropped and uncropped test sets for 11 **classes**. ResNet-50 with augmentation and discriminative learning performed the best.*
***Right:** Performance of ResNet-50 for classification over all **18 classes** of elephants.*

Augmentation	Cropped	Uncropped
No Aug.	85.256	73.932
Aug. 1	88.675	74.573
Aug. 2	89.744	74.496
Aug. 3	4.487	5.128
Aug. 4	90.812	72.863
Aug. 4 + Discrim. Learning	91.957	75.786

6. Future Work

- Extract the elephant from the image using U-Net to decrease the influence of the background
- Using an attention layer in-between pretrained ResNet-50 model to give more focus on the elphant

***Acknowledgments:** Data was provided by Dr. Daniella E. Chusyd, PhD.*