

How Can Paul Cezanne Improve the Generalizability of Deep Learning Vision Models?

Nathan Roberts | Advisor: M. R. Hasan | Electrical and Computer Engineering

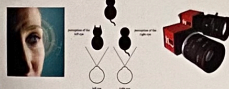


NEBRASKA RESEARCH DAYS

March 28, 2023

Introduction & Motivation

- Deep Learning (DL) vision models solve perceptual tasks (e.g., recognizing and distinguishing object types) by learning internal representations of objects from raw pixels in photographs or videos.
- However, the captured perspective of the camera does not necessarily align with how humans actually perceive the world. For example,
 - unlike the single-point perspectives in photographs, human perspectives are often **skewed** and **multi-point**.
 - the depth perception is not actually visible to humans but is constructed by the human eye.



Research Problem

- How do we enable DL models to acquire human-like perceptions?

Hypothesis

- Learning human-like visual perceptions will improve a DL model's ability to generalize.

Approach

- Instead of using photographs of objects, we train a DL model using augmented versions of the photographs that capture human-like perspectives of the objects.
- We create augmented images using the style of the late 19th-century painter **Paul Cezanne**.

Methodology

- Augment images in the style of Cezanne by using a style-transfer DL model.
- Train a DL convolutional neural network (CNN) model with the augmented images.
- Compare its performance with a baseline DL model trained on original images.



Main Contribution

Augmenting reality (input data) in the style of Cezanne's paintings exhibits the promise of improving the learning of a DL model's internal representations and its decision-making process.

Experimental Setup

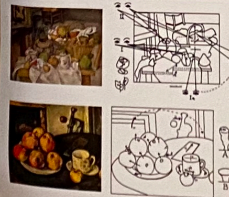
- DL Style-transfer Model: CycleGAN
- DL Classification Model: VGG-19 (Visual Geometry Group)
- Interpretability Method: Score-CAM class activation mapping
- Original Dataset: Flowers (Daisy, Rose, Dandelion, Sunflower, and Tulip)
- Augmented Datasets: Still Life With Fruit, Still Life, and People



An image of a dandelion augmented in the following styles of Cezanne's paintings: Still Life With Fruit, Still Life, and People

Why Cezanne?

- Cezanne's paintings represent **human-like perceptions** of objects. E.g.,
 - depict objects in a "lived perspective", i.e., show an object from more than a single point of view.
 - use repetitive brushstrokes to delineate each object and the relationship of one object to another.



Scientific Research Question # 1

- Do the images augmented with Cezanne's style improve a DL model's generalizability?

Flowers-Original																															
Test Accuracy	0.793																														
Classification Report	<table border="1"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>F1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.75</td> <td>0.78</td> <td>0.76</td> <td>160</td> </tr> <tr> <td>1</td> <td>0.84</td> <td>0.78</td> <td>0.81</td> <td>128</td> </tr> <tr> <td>2</td> <td>0.86</td> <td>0.86</td> <td>0.86</td> <td>140</td> </tr> <tr> <td>3</td> <td>0.82</td> <td>0.85</td> <td>0.83</td> <td>140</td> </tr> <tr> <td>4</td> <td>0.82</td> <td>0.82</td> <td>0.82</td> <td>140</td> </tr> </tbody> </table>		precision	recall	F1-score	support	0	0.75	0.78	0.76	160	1	0.84	0.78	0.81	128	2	0.86	0.86	0.86	140	3	0.82	0.85	0.83	140	4	0.82	0.82	0.82	140
	precision	recall	F1-score	support																											
0	0.75	0.78	0.76	160																											
1	0.84	0.78	0.81	128																											
2	0.86	0.86	0.86	140																											
3	0.82	0.85	0.83	140																											
4	0.82	0.82	0.82	140																											

Performance of the VGG model trained on the original images

Cezanne-People		Cezanne-Still Life with Fruit		Cezanne-Still Life																																																																															
Test Accuracy	0.823	0.770	0.805																																																																																
Classification Report	<table border="1"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>F1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.76</td> <td>0.76</td> <td>0.76</td> <td>160</td> </tr> <tr> <td>1</td> <td>0.81</td> <td>0.78</td> <td>0.79</td> <td>128</td> </tr> <tr> <td>2</td> <td>0.82</td> <td>0.85</td> <td>0.83</td> <td>140</td> </tr> <tr> <td>3</td> <td>0.81</td> <td>0.86</td> <td>0.83</td> <td>140</td> </tr> </tbody> </table>		precision	recall	F1-score	support	0	0.76	0.76	0.76	160	1	0.81	0.78	0.79	128	2	0.82	0.85	0.83	140	3	0.81	0.86	0.83	140	<table border="1"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>F1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.75</td> <td>0.78</td> <td>0.76</td> <td>160</td> </tr> <tr> <td>1</td> <td>0.84</td> <td>0.78</td> <td>0.81</td> <td>128</td> </tr> <tr> <td>2</td> <td>0.86</td> <td>0.86</td> <td>0.86</td> <td>140</td> </tr> <tr> <td>3</td> <td>0.82</td> <td>0.85</td> <td>0.83</td> <td>140</td> </tr> <tr> <td>4</td> <td>0.82</td> <td>0.82</td> <td>0.82</td> <td>140</td> </tr> </tbody> </table>		precision	recall	F1-score	support	0	0.75	0.78	0.76	160	1	0.84	0.78	0.81	128	2	0.86	0.86	0.86	140	3	0.82	0.85	0.83	140	4	0.82	0.82	0.82	140	<table border="1"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>F1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.76</td> <td>0.76</td> <td>0.76</td> <td>160</td> </tr> <tr> <td>1</td> <td>0.81</td> <td>0.78</td> <td>0.79</td> <td>128</td> </tr> <tr> <td>2</td> <td>0.82</td> <td>0.85</td> <td>0.83</td> <td>140</td> </tr> <tr> <td>3</td> <td>0.81</td> <td>0.86</td> <td>0.83</td> <td>140</td> </tr> </tbody> </table>		precision	recall	F1-score	support	0	0.76	0.76	0.76	160	1	0.81	0.78	0.79	128	2	0.82	0.85	0.83	140	3	0.81	0.86	0.83	140
	precision	recall	F1-score	support																																																																															
0	0.76	0.76	0.76	160																																																																															
1	0.81	0.78	0.79	128																																																																															
2	0.82	0.85	0.83	140																																																																															
3	0.81	0.86	0.83	140																																																																															
	precision	recall	F1-score	support																																																																															
0	0.75	0.78	0.76	160																																																																															
1	0.84	0.78	0.81	128																																																																															
2	0.86	0.86	0.86	140																																																																															
3	0.82	0.85	0.83	140																																																																															
4	0.82	0.82	0.82	140																																																																															
	precision	recall	F1-score	support																																																																															
0	0.76	0.76	0.76	160																																																																															
1	0.81	0.78	0.79	128																																																																															
2	0.82	0.85	0.83	140																																																																															
3	0.81	0.86	0.83	140																																																																															

Performance of the VGG model trained on the augmented images

SRQ# 1 result: Overall, augmented images improve the generalizability of a model by at least 3%. However, it depends on the style of the Cezanne paintings used for augmentation.

Scientific Research Question # 2

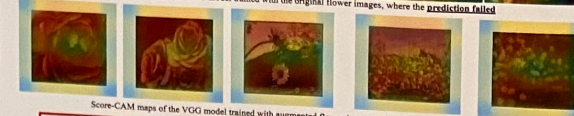
- How does the Cezanne-transformed reality (augmented images) influence a DL model's decision-making process?



Original flower images



Score-CAM maps of the VGG model trained with the original flower images, where the **prediction failed**



Score-CAM maps of the VGG model trained with augmented flower images, where the **prediction succeeded**

SRQ# 2 result: Unlike the baseline model, the augmented image-based model tends to focus on larger areas of objects and the focus often includes multiple objects enabling better decision-making.