Experiments, Results & Analysis

Addressing High Variance
Tuned Dropout, Simplified Model & Data Augmentation

![Figure 3: VanillaCNN (Baseline)](https://example.com/figure3.png)
![Figure 4: VanillaCNN-DA](https://example.com/figure4.png)

**Table 2: VanillaCNN Training and Dev Accuracy (Tuning Dropout Probability)**

<table>
<thead>
<tr>
<th>Dropout Prob</th>
<th>Learning Rate</th>
<th>Epochs</th>
<th>Train Acc (%)</th>
<th>Dev Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1e-4</td>
<td>50</td>
<td>97.16</td>
<td>73.40</td>
</tr>
<tr>
<td>0.2</td>
<td>1e-4</td>
<td>50</td>
<td>99.91</td>
<td>73.94</td>
</tr>
<tr>
<td>0.3</td>
<td>1e-4</td>
<td>50</td>
<td>99.29</td>
<td>69.68</td>
</tr>
<tr>
<td>0.4</td>
<td>1e-4</td>
<td>50</td>
<td>97.34</td>
<td>73.94</td>
</tr>
<tr>
<td>0.5</td>
<td>1e-4</td>
<td>50</td>
<td>98.94</td>
<td>69.68</td>
</tr>
<tr>
<td>0.6</td>
<td>1e-4</td>
<td>50</td>
<td>93.97</td>
<td>74.47</td>
</tr>
<tr>
<td>0.7</td>
<td>1e-4</td>
<td>50</td>
<td>96.93</td>
<td>77.00</td>
</tr>
<tr>
<td>0.8</td>
<td>1e-4</td>
<td>50</td>
<td>99.11</td>
<td>70.74</td>
</tr>
</tbody>
</table>

**Table 6: Final Tuned Model Train and Test Accuracies**

<table>
<thead>
<tr>
<th>Model</th>
<th>Dropout Rate</th>
<th>Epochs</th>
<th>Train Acc (%)</th>
<th>Test Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VanillaCNN (Baseline)</td>
<td>1e-4</td>
<td>0.7</td>
<td>50</td>
<td>99.11</td>
</tr>
<tr>
<td>VanillaCNN-DA</td>
<td>1e-4</td>
<td>0.2</td>
<td>50</td>
<td>84.57</td>
</tr>
<tr>
<td>ShallowCNN-DA</td>
<td>1e-4</td>
<td>0.1</td>
<td>100</td>
<td>98.56</td>
</tr>
<tr>
<td>ShallowCNN-DA</td>
<td>1e-4</td>
<td>0.4</td>
<td>100</td>
<td>99.65</td>
</tr>
</tbody>
</table>

**Figure 5: All Incorrectly Classified Images for VanillaCNN-DA and ShallowCNN-DA**

1. **VanillaCNN** 50 epochs train 99.13 || dev 75.00 || test 77.13
   - Tuned Dropout Prob 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8
   - Learning Rate 1e-5, 3e-5, 1e-4, 3e-4, 1e-3
   - Problem: High variance
2. **ShallowCNN** 50 epochs train 84.57 || dev 76.60 || test 78.19
3. **VanillaCNN-DA** 100 epochs train 99.56 || dev 93.09 || test 93.09
4. **ShallowCNN-DA** 100 epochs train 84.57 || dev 92.02 || test 92.55

**Data Augmentation**
Vertically flip all images (double count) Randomly split 60/20/20 train/dev/test Use new train and old dev/test sets

**Tuned Dropout Prob** 0.1, 0.2, 0.3, 0.4, 0.5

**Problem: High variance**
Still large gap b/t train and dev loss, <80% acc

**Conclusion & Next Steps**

**Standard CNN architectures work!**
By tackling the issue of high variance, even standard CNN architectures are successful at distinguishing professional and amateur artwork.

**More Advanced CNNs**
- Can more advanced CNNs correctly label easy-to-misclassify images? (e.g. when simple monochrome pieces are professional vs. amateur)

**Data**
- Data augmentation by vertical flips drastically boosted accuracy!
- Can performing more data augmentation (e.g. crops, horizontal flips) further improve accuracy?
- Increasing data overall (train/dev/test)

**What are humans missing?**
- What exactly are the CNNs learning that humans find difficult to learn?
- Which features contribute most to their predictions? Color? Types of edges?

**Content vs. Style**
- Can focusing on context vs. style of images change how images are classified?
- Can attention mechanisms be applied to focus on the most distinguishing aspects?

**Acknowledgements**
I would like to thank Huizi Mao for mentoring me throughout this project and the CS230 teaching staff for the knowledge I’ve gained through this class.

**References**
4. Angela Hawley-Dolan and Ellen Winner. Seeing the mind behind the art: People can distinguish abstract expressionist paintings from highly similar paintings by children, chimps, monkeys, and elephants. Association for Psychological Science, 2011.