Question: How can we make adding lots of dice together easier using deep learning?
Input: Pictures of 5-10 dice rolled in a dice tray
Output: The sum of the values shown on the dice

266 images containing 1855 dice
8 types: d4, d6, d8, d10, d12, d20, d100
Segmented into individual die crops using YOLO900011 (CS 231A: Computer Vision)
Hand-labeled with ground truth (value on die)
Processing: crop to square, downsize to 128x128, convert to grayscale, mean subtraction
*Augmentation: rotate 90°, 180°, 270°; randomly generate 4 square crops per rotation (min 60% original size, max 140%, roughly centered); 16x increase in training set size

CNN with softmax output (k=20 classes)
Multiclass cross-entropy loss:
\[ L(y, \hat{y}) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} y_i \log \hat{y}_j + \lambda \|W\|^2 \]
Experimented w/ different hyperparameters for batch norm (ε), dropout (p), and L2 regularization (λ)
Kept constant learning rate α=0.001

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3 conv layers, 1 FC layer, no BN, p=1.0, λ=0.0, α=0.001</td>
<td>0.537464</td>
<td>31.14%</td>
<td>12.43%</td>
</tr>
<tr>
<td>2</td>
<td>3 conv layers, 1 FC layer, BN p=0.9, p=1.0, λ=0.0, α=0.001</td>
<td>0.575043</td>
<td>39.12%</td>
<td>17.30%</td>
</tr>
<tr>
<td>3</td>
<td>3 conv layers, 1 FC layer, BN w/ ε=0.9, p=0.5, λ=0.0, α=0.001</td>
<td>0.702323</td>
<td>39.19%</td>
<td>18.92%</td>
</tr>
<tr>
<td>4</td>
<td>3 conv layers, 1 FC layer, w/ ε=0.9, p=0.5, λ=0.01, α=0.001</td>
<td>1.087315</td>
<td>40.46%</td>
<td>23.24%</td>
</tr>
</tbody>
</table>

Trained on 50,000 batches only.

Future Work
- Keep training! 50,000 batches isn’t enough.
- After tackling bias by training longer and/or increasing # of layers, tackle variance by increasing regularizations (p and λ).
- Remove problematic die types from dataset and verify that accuracy increases.

Discussion
Task is difficult, especially with certain die types.
Uneven distribution of classes across dataset:
- 0-4: 62% of dataset
- 5-9: 29% of dataset
- 10-19: 9% of dataset
Issues with bias and variance!

A picture of a d20 is sometimes difficult even for a human to label: what value is shown on this die? 16 or 8?
d4s are pyramid-shaped and don’t have an upward-facing side. Network needs to handle this die type differently.