Problem
Nature Conservancy, an NGO, monitors illegal fishing by installing cameras on fishing boats. Image classification using computers can help them to speed up and automate the process. We built a multiclass classification system for 8 different species.

Data
- **Original dataset**: 3777 images, 8 labels
- **Data preprocessing and augmentation**:
  - Feature-wise centering set input mean to 0
  - Random rotations, shifts, flips increased dataset size to >20,000 images
  - Rescaled to 224x224 for easier processing
- **Challenge**: fish are tiny part of most images

Baseline models
- Naïve Bayes (by occurrences of colors): 8%
- 1-nearest neighbor: 97%
- But this is because similar boats catch similar fish. Need more complex model for greater generalizability.

Feature Extraction
1. Scale-Invariant Feature Transform (SIFT)
   - Identifies keypoints (shapes with directional orientations)
   - Invariant to rotations and translations
   - Dense SIFT (dSIFT) runs SIFT on a grid

2. VGG-16 Convolutional Neural Network
   - 16-layer CNN, has won image recognition competitions by ImageNet
   - Pre-trained weights, outputs 1000 features

3. CIFAR-10 Convolutional Neural Network
   - 10 convolutional layers
   - Trained our own weights

Results
<table>
<thead>
<tr>
<th>Extraction Method</th>
<th>Model</th>
<th>Training Acc.</th>
<th>Test Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG-16</td>
<td>RF</td>
<td>99.5%</td>
<td>83.4%</td>
</tr>
<tr>
<td></td>
<td>SVM-RBF</td>
<td>44.3%</td>
<td>45.6%</td>
</tr>
<tr>
<td></td>
<td>SVM-Linear</td>
<td>54.1%</td>
<td>53.3%</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>-</td>
<td>95.6%</td>
<td>77.3%</td>
</tr>
</tbody>
</table>

Conclusions
- VGG-16 to RF performed best
- However, difficult to extract most important features from VGG-16
- Better and more generalizable results can be obtained by extracting fish from images, possibly using another CNN