Food images dominate across social media platforms, driving the restaurant and travel industries, but are still relatively unorganized. The ability to properly label/classify food images could lead to better recommendation systems (matching food based on an individual’s tastes and preferences, or diet). Input is a food image, output is label prediction by CNN (whether trained for scratch or pretrained).

**Data and Features**

- Food-101\(^1\): Total of 101,000 images from 101 distinct classes of food, with 1000 images per class. Of these 750 are training images that may be noisy or even mislabeled; 250 are correctly labeled validation images.
- ImageNet\(^2\): Used only during transfer learning as pre-trained weights, not directly. Of 1000 classes, 10 are food-related.
- Images are color-normalized and augmented through scaling, rotation, flipping, etc.

**Introduction**

- Food images dominate across social media platforms, driving the restaurant and travel industries, but are still relatively unorganized.
- The ability to properly label/classify food images could lead to better recommendation systems (matching food based on an individual’s tastes and preferences, or diet).
- Input is a food image, output is label prediction by CNN (whether trained for scratch or pretrained).

**Baseline model was a shallow CNN with filters of the same size, then a fully-connected layer.**

**Transfer learning was performed with VGG16, ResNet50, and InceptionV3, with top layer removed and retrained on the 101 food classes. More layers were incrementally unfrozen to improve performance.**

**Loss function was categorical cross-entropy loss:**

\[
L(y, \hat{y}) = - \sum_{c=1}^{M} y_c \log(\hat{y}_c)
\]

**Results**

- On smaller models, issue was underfitting. With larger models, issue was generally overfitting, though data augmentation and model structure (e.g. residual blocks in ResNet50) mitigated this.
- Highest accuracy model was InceptionV3 pre-trained on ImageNet and with top few layers made trainable, with 61.4% top-1 and 85.2% top-5 accuracy.
- Higher top-5 accuracy shows that model is confused by visually-similar food types.

**Future**

- Hyperparameter search and optimization.
- Bounding box image preprocessing model.
- Train separate models for food sub-categories.

**References**