CRIMSON NUTRITION

Authors

Dhruvil Joshi, Arju Singh

Can computer vision accurately identify food items from a regular image taken by the user?

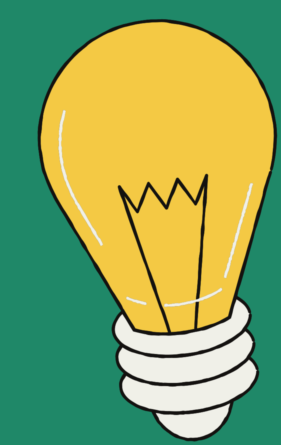
MOTIVATION

With growing health consciousness, we aimed to explore the intersection of food, nutrition, and technology.

Work towards helping people make healthier choices by understanding the nutritional value and quality of the food they consume.

OBJECTIVE

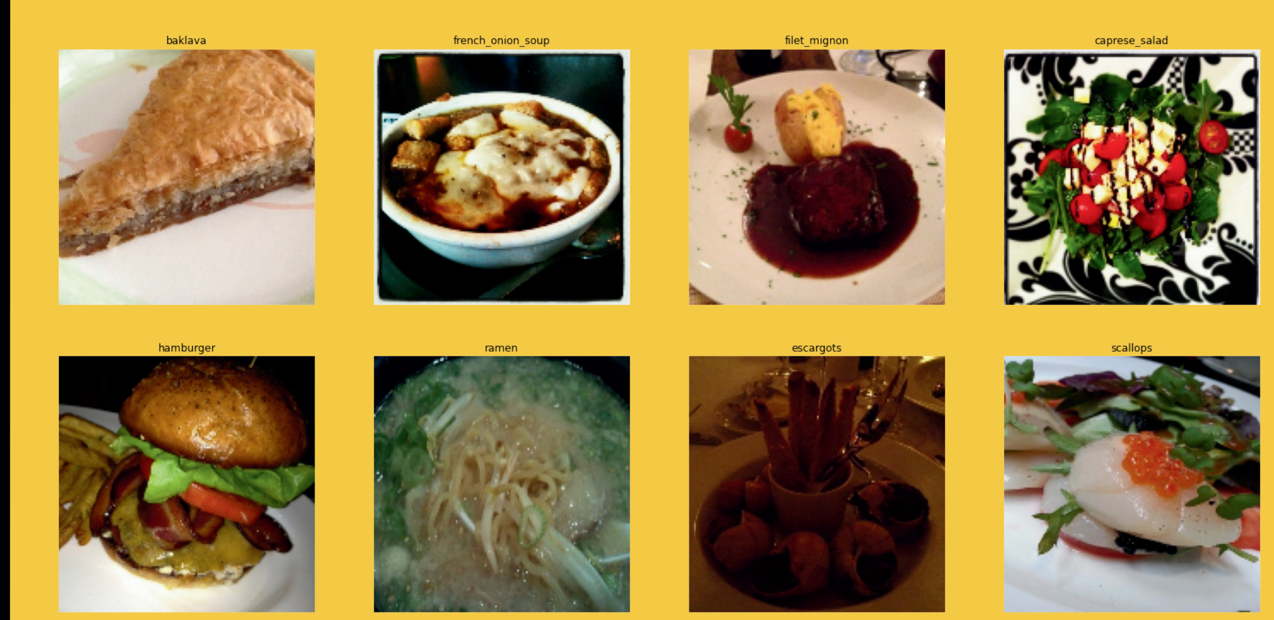
Work on identifying which food item from the image uploaded by the user and providing it's nutritional values.



SETUP

DATASET:

- Used the Food101 dataset with 101 food categories and 1,000 images per class.
- All images resized to 224×224 pixels to fit ResNet input requirements.
- Applied data augmentation: random crops, flips, rotations, color jittering, and ImageNet normalization.
- Created a custom nutritional dataset by mapping values to the 101 food classes.



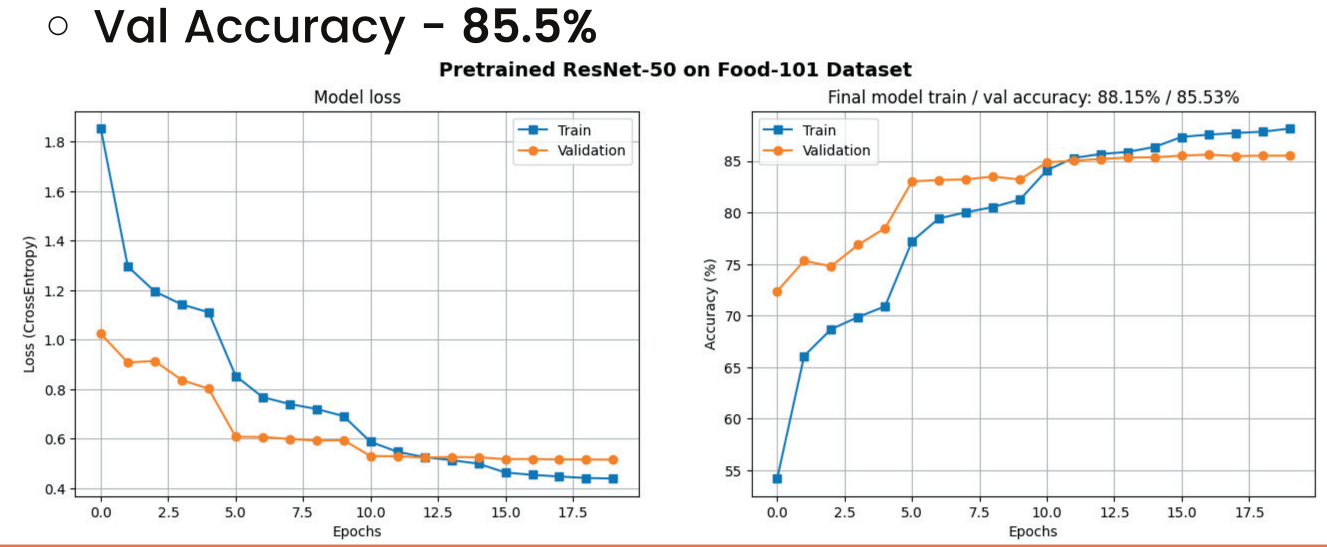
TECHNICAL:

- Chose ResNet-50 over MobileNetV2 for better feature extraction and deeper architecture.
- Integrated Groq's LLaMA-3.3-70B (free tier) LLM to generate instant nutritional insights for predicted food items which can help the user.

TRAINING DETAILS

We used the pretrained ResNet-50 model and fine tuned the final layers on the dataset.

- **Hyperparameters**:
  - Dataset Split: 75% train/25% val
  - Learning rate: 0.001
  - L2 regularization: 1e-3
  - Batch size: 256
  - Training epochs: 20
- **Optimization technique**:
  - Adam optimizer with weight decay
  - StepLR scheduler (step_size=5, gamma=0.2)
  - Cross-entropy loss function
- **Results**:
  - Train Accuracy - **88.15 %**
  - Val Accuracy - **85.5%**

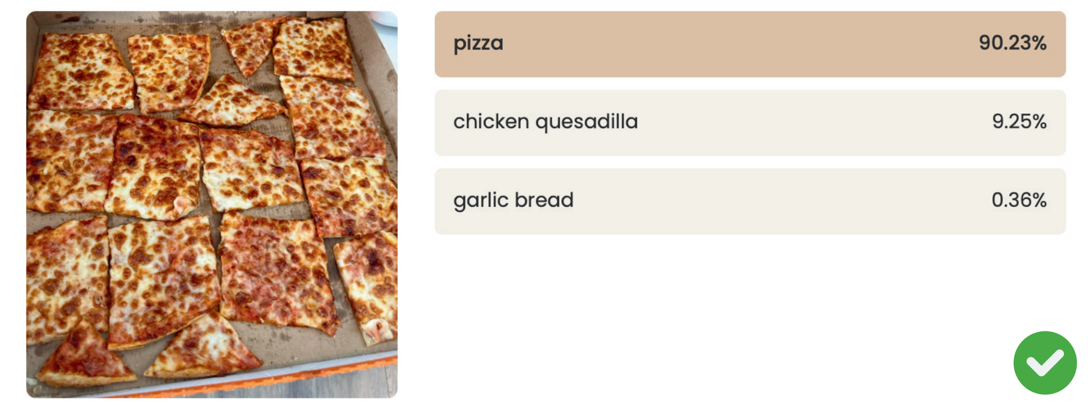


FUTURE SCOPE

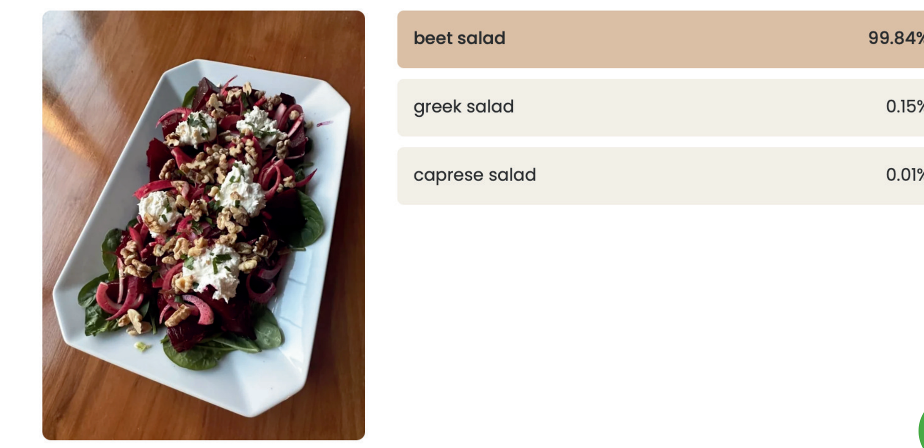
1. Extend to multi-item detection for identifying all foods in a single plate or image
2. Explore calorie estimation and ingredient-level analysis for deeper nutritional insights
3. Integrating it with IU Dining and Hospitality NetNutrition®

RESULTS/FINDINGS

1. Foods with unique visual signatures achieve higher classification accuracy



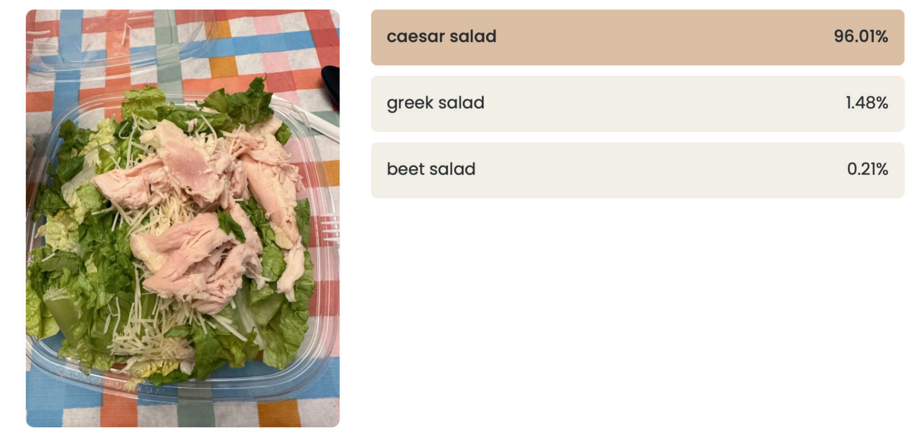
Pizza seems to be correctly classified, in this example the pizza isn't round or standard slice shaped but still model correctly classifies it.



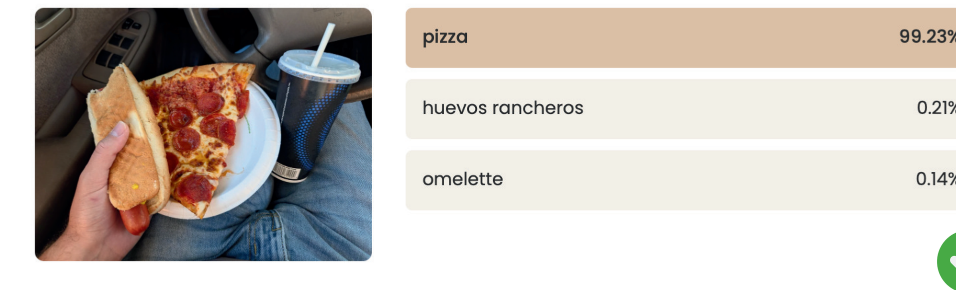
Beet root salad: the salad like features along with the red color could be the clue for the model.

2. What about multiple food items in a single photo

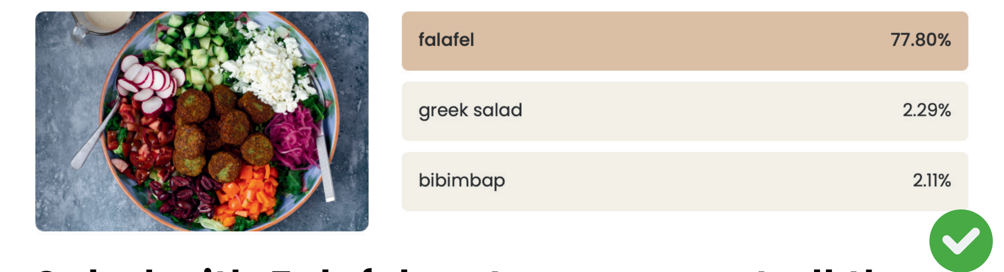
In case of such when multiple food items are there in the image, We observed that the model captures the most significant food item generally.



Caesar Salad with raw chicken on top, but the model correctly identifies the salad.



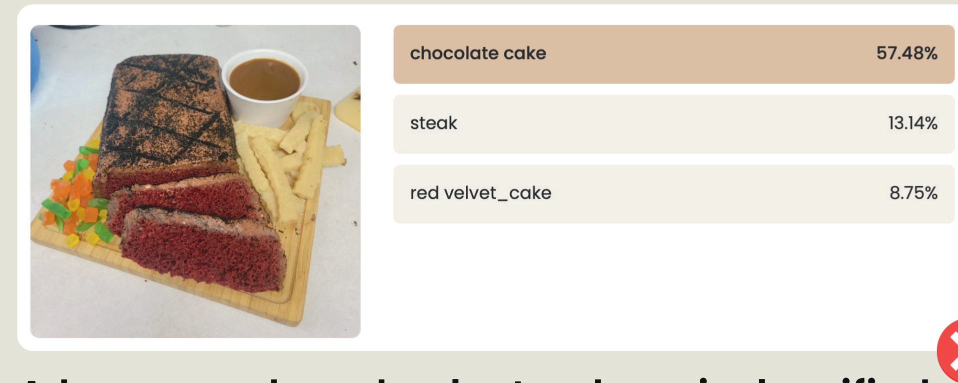
Hot dog and pizza together and the model identifies the pizza



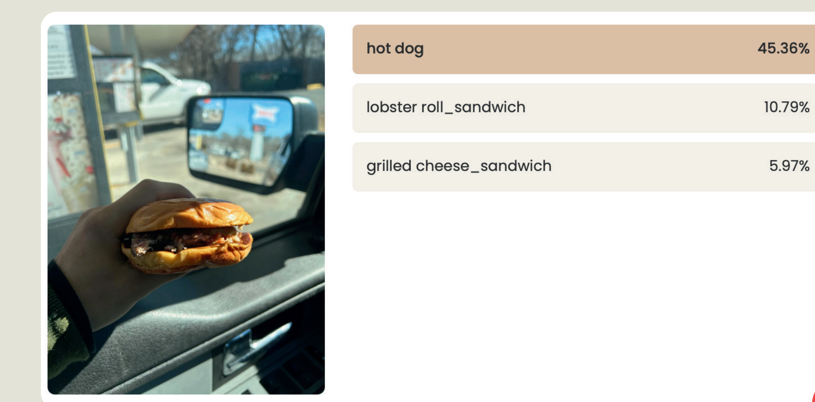
Salad with Falafel on top, amongst all the model identifies the falafels

CHALLENGES

1. Similar looking food items often confuse the model and it fails many times there



A homemade red velvet cake misclassified as chocolate cake (likely due to dark exterior), then steak (texture), and finally red velvet cake.



A pulled pork sandwich misclassified as hot dog, but the predictions are reasonable and not entirely off.

2. Only 101 food items for now

We trained our model only on the 101 classes present in the dataset, there are clearly many more categories of food in the real world than just 101 which our model has. This limitation restricts the application's coverage, particularly for regional and ethnic cuisines that are underrepresented in the Food101 dataset.

CONCLUSION

We explored food and nutrition using computer vision by trying CNN based ResNet-50, MobileNetV2 models on the Food101 dataset, achieving decent validation accuracy. While results were promising, issues like similar-looking foods and limited class variety leave room for improvement.

References

1. Amugongo, et al. Mobile Computer Vision-Based Applications for Food Recognition and Volume and Calorific Estimation (2023)
2. Tahir, et al. A Comprehensive Survey of Image-Based Food Recognition & Volume Estimation Methods for Dietary Assessment (2023)