Data-centred Deep Learning Models for Food Fine-grained recognition

Petia Radeva
Full professor
Head of “Artificial Intelligence and Biomedical Applications”
Consolidated Research Group,
Universitat de Barcelona, Spain

• Joint work with Eduardo Aguilar, Bhalaji Nagarajan, Imanol Gonzalez, Jesus Molina, Ricardo Marquez, etc
1. Why Food Recognition?

2. Self-Supervised Learning for Fine-Grained Recognition
   - Validation of All4One

3. Other Food Recognition works

4. Food Recognition real applications
## What are the most popular datasets today?

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Papers</th>
<th>Benchmarks</th>
<th>Images (K)</th>
<th>Classes</th>
<th>Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cifar-10</td>
<td>10581</td>
<td>66</td>
<td>60</td>
<td>10</td>
<td>32x32</td>
</tr>
<tr>
<td><strong>ImageNet</strong></td>
<td>10046</td>
<td>97</td>
<td><strong>1400</strong></td>
<td><strong>1000</strong></td>
<td>variable</td>
</tr>
<tr>
<td>COCO</td>
<td>7160</td>
<td>78</td>
<td>123</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>OpenImages</td>
<td></td>
<td></td>
<td>9000</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>MNIST</td>
<td>5911</td>
<td>49</td>
<td>60</td>
<td>10</td>
<td>28x28</td>
</tr>
<tr>
<td>Cifar-100</td>
<td>5322</td>
<td>42</td>
<td>60</td>
<td>100</td>
<td>32x32</td>
</tr>
<tr>
<td>Cityskapes</td>
<td>2562</td>
<td>37</td>
<td>25</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>SVHN</td>
<td>2474</td>
<td>11</td>
<td>60</td>
<td>10</td>
<td>32x32</td>
</tr>
<tr>
<td>Kitti</td>
<td>2453</td>
<td>120</td>
<td>0,5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>CelebA</td>
<td>2408</td>
<td>20</td>
<td>202</td>
<td>10177</td>
<td>178x218</td>
</tr>
<tr>
<td>Fashion-MNIST</td>
<td>2150</td>
<td>17</td>
<td>70</td>
<td>10</td>
<td>28x28</td>
</tr>
<tr>
<td>CUB-00-2011</td>
<td>2408</td>
<td>37</td>
<td>12</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td><strong>Places</strong></td>
<td>760</td>
<td>4</td>
<td><strong>2500</strong></td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>Tiny ImageNet</td>
<td>516</td>
<td>7</td>
<td>31</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td><strong>Places205</strong></td>
<td>468</td>
<td>1</td>
<td><strong>2500</strong></td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>Caltech-101</td>
<td>393</td>
<td>6</td>
<td>5</td>
<td>101</td>
<td>300x200</td>
</tr>
<tr>
<td>Stanford Cars</td>
<td>392</td>
<td>8</td>
<td>16</td>
<td>196</td>
<td>360x240</td>
</tr>
<tr>
<td>Caltech-256</td>
<td>345</td>
<td>4</td>
<td>30</td>
<td>257</td>
<td></td>
</tr>
</tbody>
</table>

Large Scale Food Recognition Dataset

Weiqing Min, Zhiheng Wang, Yuxin Liu, Mengjiang Luo, Liping Kang, Xiaomei Wei, Xia... All Authors

Abstract: Food recognition plays an important role in food choice and intake, which is essential to the health and well-being of humans. It is thus of importance to the computer vision community, and can further support many food-oriented vision and multimodal tasks, e.g., food detection and segmentation, cross-modal recipe retrieval and generation.

More than 1M images!

Food recognition popularity

Number of Food recognition papers

iFood 2011 fine-grained (prepared) food categories with 135733

AICrowd: 26000 annotated segmented images

LargeFineFoodAI: 1,000 fine-grained food categories and over 50,000 images.
Food image analysis
Why is the food recognition a challenge?
Motivation

Food Analysis Problems

Ingredients

- Intra-class variability
- Inter-class similarity

Decreasing in Precision

Intra-class variability example: Apple. Image source: Recipes5k

Inter-class similarity example: Tomato sauce and Curry sauce. Image source: Recipes5k
The food recognition is a Fine-grained recognition problem
Challenges of Food image analysis

Food256: 25,600 images (100 images/class) Classes: 256

Food101 – 101,000 images (1000 images/class) Classes: 101

FoodX-251
Classes: 251
140K images

Food1K
Classes: 1000
370K images

Food DB
150,000 images
231 categories

ImageNet
1,400,000 images
1000 categories

Future Food DB
?????? images
200,000 categories

Current SoA on Food recognition
- 79% on UECFOOD
- 44% on ChinaFood1000
How to leverage from the huge amount of non-annotated data/images?

Self-Supervised Learning allows to leverage big amounts of unlabelled data to make NNs more robust!
Complex Problems Need Powerful Models

What makes this project possible?

Category: noodles and pasta
Dish type: spaghetti carbonara

Ingredients:
eggs, olive oil, garlic, black pepper, white wine, parmesan cheese, spaghetti, pancetta, grated pecorino

Nutritional information:
Energy: 710.7 kcal
Sugars: 3.7 gr
Cholesterol: 176.9 mg
1. **Transfer Learning**: The use of pre-trained models and transfer learning techniques was becoming more widespread, as they can significantly reduce the amount of data required to train effective models.
   - Multi-task learning

2. **Uncertainty modeling**: refers to the process of quantifying and managing uncertainty or ambiguity in the predictions or decisions made by machine learning models
   - Uncertainty-aware MTL
   - Learning with noisy labeling

3. **Self-Supervised Learning**: methods train models on unlabeled data, which can be particularly useful in cases where labeled data is scarce.

4. **Meta-Learning**: Meta-learning techniques were explored to enable models to learn how to learn, which can lead to faster adaptation to new tasks.
   - Continual learning

5. **Generative AI and Generative Adversarial Networks (GANs)**: GANs were being used for a variety of applications, from image generation to data augmentation and domain adaptation.
   - Uncertainty-aware data augmentation
   - NeRFs
   - Stable diffusion
Trends in DL

• **6. Efficiency and Model Compression**: There was a growing interest in making deep learning models smaller, faster, and more energy-efficient, particularly for edge computing and mobile applications.
  • Scaling by hierarchical DL models
  • Fine-grained recognition

• **7. Multimodal Learning**: Combining information from different sources, such as text and images, to create more comprehensive models for understanding and generating content.
  • Food ontology

• **8. Explainable AI (XAI)**: The need for understanding and interpreting deep learning models became more critical, especially in fields like healthcare.
  • Robust Explainable models

• **9. Federated Learning**: This approach allows for training models across decentralized data sources without sharing raw data, preserving privacy. It gained traction, particularly in applications involving sensitive or proprietary data.

• **10. AI for Healthcare**: Deep learning was increasingly applied to medical image analysis, drug discovery, and patient diagnosis, with a focus on improving healthcare outcomes.
  • Our projects: food data analysis,
Data-centric Food image analysis

Large-scale Food recognition

Learning with Noisy Labeling Food recognition

Food ontology-based Deep learning

Food image analysis

Uncertainty modelling

Fine-grained Food recognition

Self-supervised Learning

Generative AI for Food Volume Estimation
1. Why Food Recognition?

2. Self-Supervised Learning for Fine-Grained Recognition

   - Validation of All4One

3. Other Food Recognition works

4. Food Recognition real applications
Neural networks are dragons, but...

Greedy dragons!
How to make a NN robust?

Use data augmentation
How to make a NN robust with unlabeled data?
What can you do if you have a lot of just data and may be a not-trained model?

Contrastive loss

\[ L_{\text{InfoNCE}}^{(z_i, z_i^+)} = - \log \frac{\exp(z_i \cdot z_i^+ / \tau)}{\exp(z_i \cdot z_i^+ / \tau) + \sum_{z \in N_i} \exp(z_i \cdot z^- / \tau)} \]
What is self-supervised learning (SSL)?

What would you do if you have thousands of unlabelled images?
Self-Supervised Learning: Benefits & Uses in 2023

Yann LeCun and Yoshua Bengio: “Self-supervised learning is the key to human-level intelligence”

https://research.aimultiple.com/self-supervised-learning/

Yann LeCun, VP and Chief AI Scientist at Facebook, is explaining how self-supervised learning works at PAISS’19.
Babies learn how the world Works by observation.

Largely by observation, with remarkably little interaction.

Photos courtesy of Emmanuel Dupoux.
Artifical vs Natural NNs

Understand brain through NNs:

- the brain is full of feedback connections, while current models have few such connections, if any.

Next step: use SSL to train highly recurrent networks and see how the activity in NNs compares to real brain activity.

Crucial step: match the activity of NNs in SSL models to the activity of individual biological neurons.

Hypothesis: the visual systems of humans and other primates are the best studied of all animal sensory systems,
- neuroscientists struggle to explain why they include two separate pathways:
  - the ventral visual stream, which is responsible for recognizing objects and faces, and
  - the dorsal visual stream, which processes movement (the “what” and “where” pathways, respectively).

“No doubt that 90% of what the brain does is self-supervised learning,” Blake Richards, a computational neuroscientist at McGill University and Mila, the Quebec Artificial Intelligence Institute.
SimCLR by the Google AI team

**Introduces projectors:** a learnable nonlinear transformation between the representation and the contrastive loss

**Positive sampling:** Given a batch of N samples, the pretext task P generates two augmented views $x^a_i$ and $x^+i$ for each sample $x_i$ of the batch.

**Negative sampling:** the rest of the images $x^-i$ on the same batch to form the negative pairs $(x^a_i, x^-i)$.

Batch sizes of 8196 are used.

**Loss function:**

$$L_i^{SimCLR} = -\log\left(\frac{\exp(z^a_i \cdot z^+/\tau)}{\sum_{k=1}^{N} \exp(z^a_i \cdot z^-_k/\tau)}\right)$$

State-of-the-art Contrastive SSL models

Negative sampling augment the cost

Only positive sampling could lead trivial solutions
SSL Framework: Barlow Twins

Self-supervised Learning

Still SSL learn too fast!

Overfitting the domain!
SSL Framework: NNCLR

\[ \mathcal{L}_{i}^{\text{NNCLR}} = -\log \frac{\exp \left( \text{NN}(z_i, Q) \cdot z_i^+ / \tau \right)}{\sum_{k=1}^{n} \exp \left( \text{NN}(z_i, Q) \cdot z_k^+ / \tau \right)} \]
Common neighbour contrastive approaches only contrast the first neighbour.

We create representations that contain contextual information from the k NNs - Contrast it in a single objective computation:

\[
L_{i}^{\text{centroid}} = -\log \left( \sum_{n=1}^{n} \frac{\exp(c_i^1 \cdot c_i^2 / \tau)}{\exp(c_i^1 \cdot c_n^2 / \tau)} \right)
\]
Self-Attention in order to obtain “context-aware” representations

Attention(\(Q, K, V\)) = \text{softmax}\left(\frac{QK^T}{\sqrt{n}}\right)V
All for One: Centroid contrast

1. Why Food Recognition?

2. Self-Supervised Learning for Fine-Grained Recognition
   - Validation of All4One

3. Other Food Recognition works

4. Food Recognition real applications
Dataset and Evaluation Metrics

Dataset

Food-101

Evaluation Metrics

SSL Model

Top 1 and top 5 accuracy using a k-NN classifier

Classifier

Overall accuracy
Variance
Entropy
Mutual Information
### Quantitative Results: CIFAR

#### CIFAR-10

<table>
<thead>
<tr>
<th>Method</th>
<th>Backbone</th>
<th>Epochs</th>
<th>Acc@1 (Online)</th>
<th>Acc@5 (Online)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYOL</td>
<td>ResNet18</td>
<td>1000</td>
<td>92.58</td>
<td>99.79</td>
</tr>
<tr>
<td>DeepCluster V2</td>
<td>ResNet18</td>
<td>1000</td>
<td>88.85</td>
<td>99.58</td>
</tr>
<tr>
<td>DINO</td>
<td>ResNet18</td>
<td>1000</td>
<td>89.52</td>
<td>99.71</td>
</tr>
<tr>
<td>MoCo V2+</td>
<td>ResNet18</td>
<td>1000</td>
<td>92.94</td>
<td>99.79</td>
</tr>
<tr>
<td>MoCo V3</td>
<td>ResNet18</td>
<td>1000</td>
<td>93.10</td>
<td>99.80</td>
</tr>
<tr>
<td>ReSSL</td>
<td>ResNet18</td>
<td>1000</td>
<td>90.63</td>
<td>99.62</td>
</tr>
<tr>
<td>SimCLR</td>
<td>ResNet18</td>
<td>1000</td>
<td>90.74</td>
<td>99.75</td>
</tr>
<tr>
<td>Siam</td>
<td>ResNet18</td>
<td>1000</td>
<td>90.51</td>
<td>99.72</td>
</tr>
<tr>
<td>SwAV</td>
<td>ResNet18</td>
<td>1000</td>
<td>89.17</td>
<td>99.68</td>
</tr>
<tr>
<td>VlbCReg</td>
<td>ResNet18</td>
<td>1000</td>
<td>91.18</td>
<td>99.74</td>
</tr>
<tr>
<td>VICReg</td>
<td>ResNet18</td>
<td>1000</td>
<td>92.07</td>
<td>99.74</td>
</tr>
<tr>
<td>W-MSE</td>
<td>ResNet18</td>
<td>1000</td>
<td>88.67</td>
<td>99.68</td>
</tr>
<tr>
<td>Barlow Twins</td>
<td>ResNet18</td>
<td>1000</td>
<td>92.10</td>
<td>99.73</td>
</tr>
<tr>
<td>NNCLR</td>
<td>ResNet18</td>
<td>1000</td>
<td>91.88</td>
<td>99.78</td>
</tr>
<tr>
<td><strong>All4One</strong></td>
<td>ResNet18</td>
<td>1000</td>
<td><strong>93.24</strong></td>
<td><strong>99.88</strong></td>
</tr>
</tbody>
</table>

#### CIFAR-100

<table>
<thead>
<tr>
<th>Method</th>
<th>Backbone</th>
<th>Epochs</th>
<th>Acc@1 (Online)</th>
<th>Acc@5 (Online)</th>
<th>k-NN Acc@1 (Online)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYOL</td>
<td>ResNet18</td>
<td>1000</td>
<td>70.46</td>
<td>91.96</td>
<td>-</td>
</tr>
<tr>
<td>DeepCluster V2</td>
<td>ResNet18</td>
<td>1000</td>
<td>63.61</td>
<td>88.09</td>
<td>-</td>
</tr>
<tr>
<td>DINO</td>
<td>ResNet18</td>
<td>1000</td>
<td>66.76</td>
<td>90.34</td>
<td>-</td>
</tr>
<tr>
<td>MoCo V2+</td>
<td>ResNet18</td>
<td>1000</td>
<td>69.89</td>
<td>91.65</td>
<td>-</td>
</tr>
<tr>
<td>MoCo V3</td>
<td>ResNet18</td>
<td>1000</td>
<td>68.83</td>
<td>90.57</td>
<td>-</td>
</tr>
<tr>
<td>ReSSL</td>
<td>ResNet18</td>
<td>1000</td>
<td>65.92</td>
<td>89.73</td>
<td>-</td>
</tr>
<tr>
<td>SimCLR</td>
<td>ResNet18</td>
<td>1000</td>
<td>65.78</td>
<td>89.04</td>
<td>-</td>
</tr>
<tr>
<td>Siam</td>
<td>ResNet18</td>
<td>1000</td>
<td>66.04</td>
<td>89.62</td>
<td>-</td>
</tr>
<tr>
<td>SwAV</td>
<td>ResNet18</td>
<td>1000</td>
<td>64.88</td>
<td>88.78</td>
<td>-</td>
</tr>
<tr>
<td>VlbCReg</td>
<td>ResNet18</td>
<td>1000</td>
<td>67.37</td>
<td>90.07</td>
<td>-</td>
</tr>
<tr>
<td>VICReg</td>
<td>ResNet18</td>
<td>1000</td>
<td>68.54</td>
<td>90.83</td>
<td>-</td>
</tr>
<tr>
<td>W-MSE</td>
<td>ResNet18</td>
<td>1000</td>
<td>61.33</td>
<td>87.26</td>
<td>-</td>
</tr>
<tr>
<td>NNCLR</td>
<td>ResNet18</td>
<td>1000</td>
<td>69.62</td>
<td>91.52</td>
<td>-</td>
</tr>
<tr>
<td>NNCLR*</td>
<td>ResNet18</td>
<td>1000</td>
<td>60.17</td>
<td>91.70</td>
<td>62.16</td>
</tr>
<tr>
<td>Barlow Twins</td>
<td>ResNet18</td>
<td>1000</td>
<td>70.90</td>
<td>91.91</td>
<td>-</td>
</tr>
<tr>
<td>Barlow Twins*</td>
<td>ResNet18</td>
<td>1000</td>
<td>71.21</td>
<td>92.46</td>
<td>63.11</td>
</tr>
<tr>
<td>MSE*</td>
<td>ResNet18</td>
<td>1000</td>
<td>67.84</td>
<td>91.64</td>
<td>63.36</td>
</tr>
<tr>
<td><strong>All4One</strong></td>
<td>ResNet18</td>
<td>1000</td>
<td><strong>72.17</strong></td>
<td><strong>93.35</strong></td>
<td><strong>64.84</strong></td>
</tr>
</tbody>
</table>

---

### Quantitative Results: ImageNet-100

<table>
<thead>
<tr>
<th>Method</th>
<th>Backbone</th>
<th>Epochs</th>
<th>Acc@1 (online)</th>
<th>Acc@5 (online)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYOL ++</td>
<td>ResNet18</td>
<td>400</td>
<td>80.16</td>
<td>95.02</td>
</tr>
<tr>
<td>DeepCluster V2</td>
<td>ResNet18</td>
<td>400</td>
<td>75.36</td>
<td>93.22</td>
</tr>
<tr>
<td>DINO</td>
<td>ResNet18</td>
<td>400</td>
<td>74.84</td>
<td>92.92</td>
</tr>
<tr>
<td>MoCo V2+ ++</td>
<td>ResNet18</td>
<td>400</td>
<td>78.20</td>
<td>95.50</td>
</tr>
<tr>
<td>MoCo V3 ++</td>
<td>ResNet18</td>
<td>400</td>
<td>80.36</td>
<td>95.18</td>
</tr>
<tr>
<td>ReSSL</td>
<td>ResNet18</td>
<td>400</td>
<td>76.92</td>
<td>94.20</td>
</tr>
<tr>
<td>SimCLR ++</td>
<td>ResNet18</td>
<td>400</td>
<td>77.64</td>
<td>94.06</td>
</tr>
<tr>
<td>SimSiam</td>
<td>ResNet18</td>
<td>400</td>
<td>74.54</td>
<td>93.16</td>
</tr>
<tr>
<td>SwAV</td>
<td>ResNet18</td>
<td>400</td>
<td>74.04</td>
<td>92.70</td>
</tr>
<tr>
<td>ViBCReg</td>
<td>ResNet18</td>
<td>400</td>
<td>79.86</td>
<td>94.98</td>
</tr>
<tr>
<td>VICReg ++</td>
<td>ResNet18</td>
<td>400</td>
<td>79.22</td>
<td>95.06</td>
</tr>
<tr>
<td>W-MSE</td>
<td>ResNet18</td>
<td>400</td>
<td>67.60</td>
<td>90.94</td>
</tr>
<tr>
<td>Barlow Twins ++</td>
<td>ResNet18</td>
<td>400</td>
<td>80.38</td>
<td>95.28</td>
</tr>
<tr>
<td>NNCLR ++</td>
<td>ResNet18</td>
<td>400</td>
<td>79.80</td>
<td>95.28</td>
</tr>
<tr>
<td>All4One (Ours)</td>
<td>ResNet18</td>
<td>400</td>
<td>81.93</td>
<td>96.23</td>
</tr>
</tbody>
</table>
Qualitative Analysis: UMAP

Epoch 1

Epoch 100

NNCLR

Musketeer
Qualitative Analysis: NN Retrieval

1. Why Food Recognition?

2. Self-Supervised Learning for Fine-Grained Recognition
   - Validation of All4One

3. Other Food Recognition works

4. Food Recognition real applications
How to make a NN robust?

Use data augmentation
Is it enough?
Synthetic image generation
Using synthetic data but how?

- Some images could be more beneficial for the classification than others

More emphasis should be placed on these data

- **Hypothesis**: Estimating the uncertainty can help us decide the most appropriate samples and classes to perform additional data augmentation methods.

Use the data augmentation applied class-conditionally to improve the results in terms of accuracy and also to reduce the overall epistemic uncertainty.

During the prediction phase, the same image is fed to the CNN several times to calculate the epistemic uncertainty given by the model for that image.
Obtaining lexical embeddings

- Use the **class labels** from the dataset directly, pre-process them and **pass** them to the **lexical encoder**.

- **Leverage LLMs** to obtain a **list of ingredients** for a given class name.
  - The class name is used as a prompt to **generate the list of ingredients**.
  - A text composed of the class name and comma-separated ingredients is used as the textual representation of each class, which is **the input to the lexical encoder**.

- Feed with the **food images** and a **textual prompt** asking for the visible ingredients in the image.
  - A querying transformer is used to **query the input (image)** using the prompt.
  - This enables the model to produce a **list of ingredients**, in this case only the visible ones, which is used as a textual representation of each image.
  - The **captions** are produced per image, obtaining several **lists of ingredients for each class**, which are **further processed** through the lexical encoder.

- The textual representations are processed by a **neural lexical encoder** which transforms the input sentence into a fixed-length embedding in the lexical feature space.

- A **clustering** is applied to detect similar classes.

---

Trained through an end-to-end **multi-task learning** process, this method **enhances performance** in the fine-grained food recognition task, showing exceptional prowess with highly similar classes.
## Validation

Comparison of DoD with SoTA methods in Food-101

† =bigger image size. °◊ =subset-based method

<table>
<thead>
<tr>
<th>Method</th>
<th>Test Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grafit [61] (ICCV’21)</td>
<td>93.7</td>
</tr>
<tr>
<td>EffNet-B7 [59] (ICML’19) †</td>
<td>93.0</td>
</tr>
<tr>
<td>PMG [8] (CVPR’21) †$</td>
<td>87.5</td>
</tr>
<tr>
<td>FGFR [53] (Madima’22)$</td>
<td>93.8</td>
</tr>
<tr>
<td><strong>DoD + SwinV2-B$</strong></td>
<td><strong>94.9</strong></td>
</tr>
</tbody>
</table>

UMAP of the baseline and DoD
Learning with Noisy Labeling

Did you know that:

- 3.4% average error rate across all datasets,
- 6% for ImageNet
- MNIST digits dataset contains 15 (human-validated) label errors in the test set.

Towards Volume Estimation: MomentsNeRF
1. Why Food Recognition?

2. Self-Supervised Learning for Fine-Grained Recognition
   - Validation of All4One

3. Other Food Recognition works

4. Food Recognition real applications
SUCCESS STORY: FOOD intake monitoring of kidney transplant patients

LogMeal is a HealthApp and API in the cloud that is able to automatically recognize and analyze food from images.

- **Automatic food diary** construction (UB).
- **Accurate, objective and continuous** food intake monitoring (UB).
- Semi-automatic **volume estimation** (Nestle).
- **Meal planner** and **health recommendations** (Nestle).
SUCCESS STORY Greenhabit: a serious game to promote change behaviour

Ted Talk of Chantal Linders: “Manage the Monster in Your Head”

Greenhabit (EIT Digital, 2020/21)
Food recognition is a perfect test domain for powerful Machine/Deep Learning models.

Food Image Analysis is highly underexplored problem that could convert in an important benchmark for CV algorithms.

ALI4One combines its neighbour contrast objective with a feature redundancy reduction objective, being beneficial in its overall performance.

- consistently outperforms SoTA instance discrimination frameworks on popular image classification benchmarking datasets, namely, CIFAR-10, CIFAR-100 and ImageNet-100.

Multiple other CV problems are highly relevant for food image analysis: uncertainty modelling, multi-task learning, vision-language models, fine-grained recognition, multi-scale classification, etc.

Multiple real applications and professional opportunities
Our small group
Thank you!

petia.ivanova@ub.edu