

# CNN-based Fine-Grained Categorization for Food Images

G16 Chen Cai, Yuxuan Zheng, Jin Wang, Chengwen Zuo

Motivation

Dataset Construction

CNN Model

Evaluation

Conclusion

Contribution

# Motivation



- Food recognition is an important task for automatic or semi-automatic daily dietary monitoring, image retrieval, recommendation and restaurant & beverage service.
- However, in contrast to general object recognition, food images typically do not exhibit distinctive spatial arrangement and common semantic patterns.
- Fine-grained classification for food images is needed.



Insignificant  
differences  
between groups

Significant  
differences in a  
group

- Lack of large-scale food datasets
- The data from the crawler contains noise



# Dataset Construction

- Dataset contains:

- Training data: 75,296 images (naming format: LabelID\_ImageID.jpg, e.g. 0\_0000.jpg, 0\_0001.jpg, 999\_0010.jpg)
- Validation data: 10,000 images (naming format: LabelID\_ImageID.jpg, e.g. 0\_0101.jpg, 0\_0103.jpg, 999\_0116.jpg)
- Testing data: 20,000 images (without labels) (naming format: test\_ImageID.jpg, e.g. test\_0004.jpg, test\_0005.jpg)

P.S. The shape of all pictures is arbitrary, i.e. we need to resize the images.

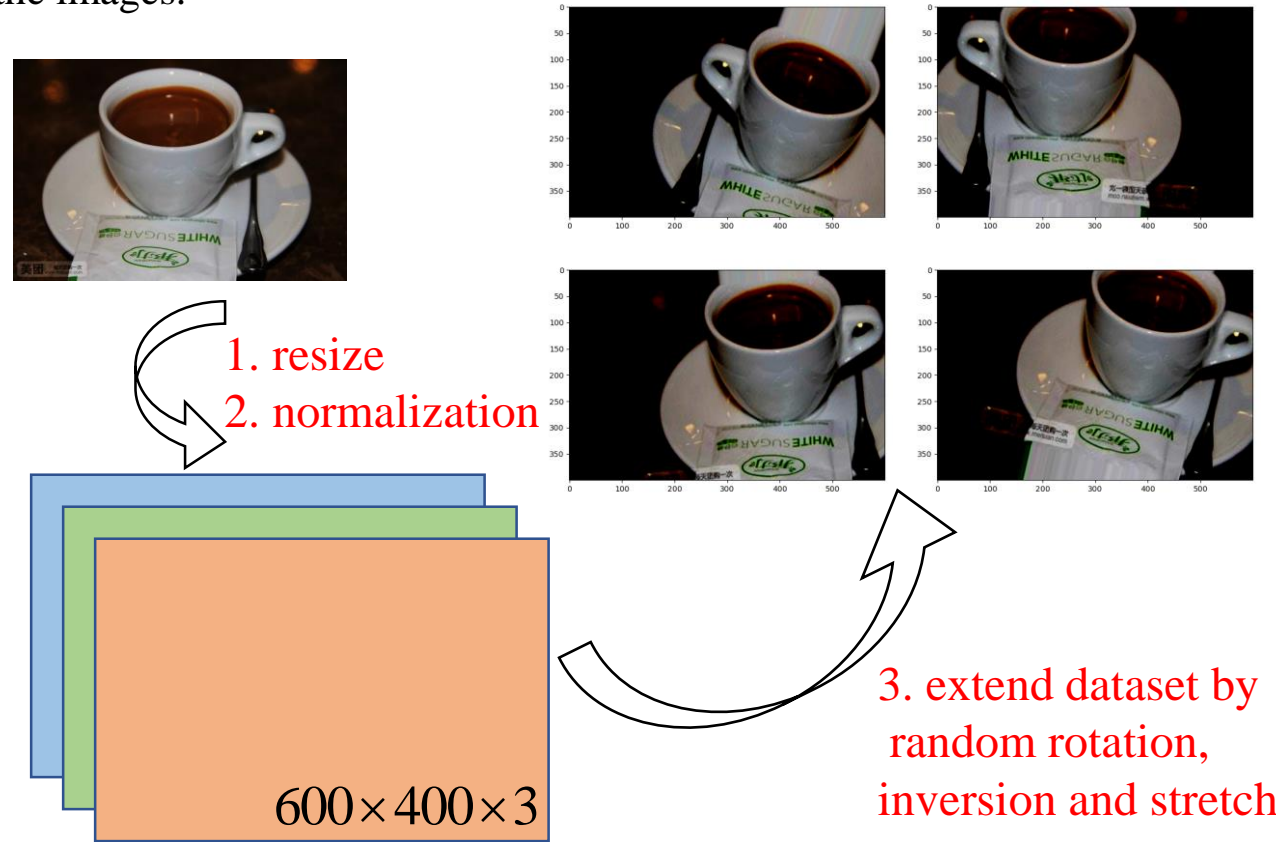
75,296 samples  
↓ ×4  
301,184 samples

- Pretreatment:

- Resize (training data, validation data, testing data)  
Since most of the images in the dataset are wide images, in order to bring the data into the model for training, all the images are changed to a size of 600 \* 400.
- Normalization (training data, validation data, testing data)  
Normalize the pixel value of the image to - 1 to 1 to improve the convergence speed and accuracy of the model.

$$\text{pixel} = 2 \times \left( \frac{\text{pixel}}{255} - 0.5 \right) = \frac{\text{pixel}}{127.5} - 1$$

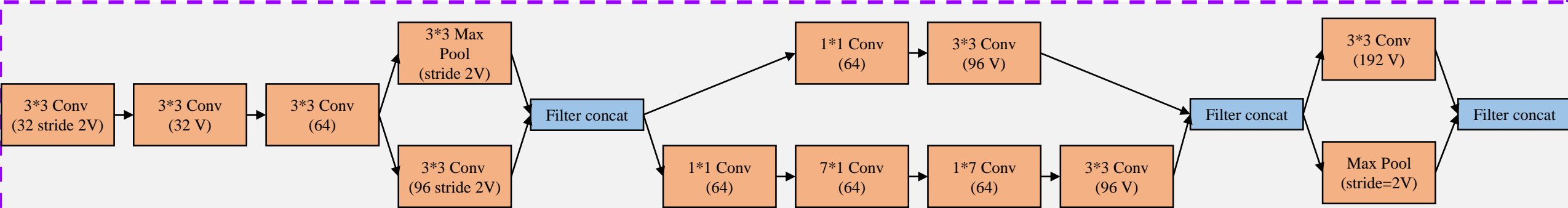
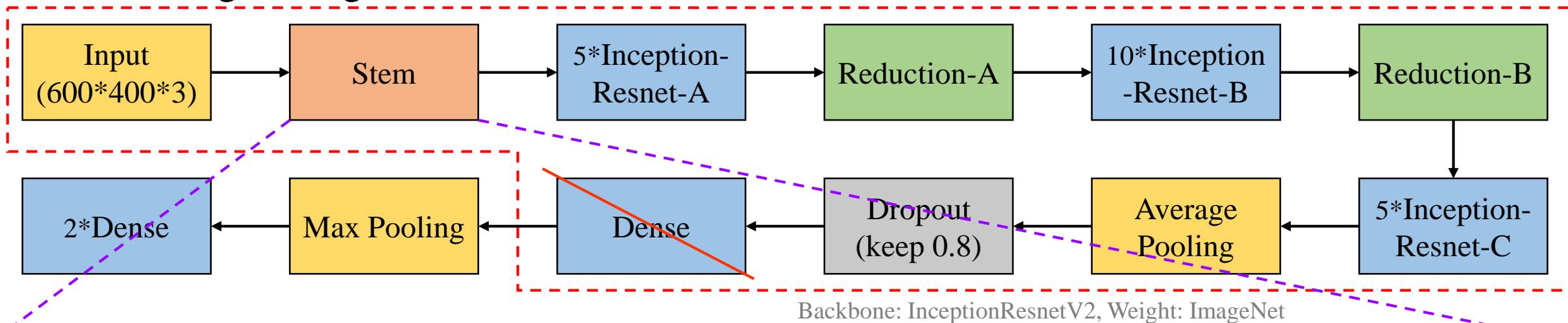
- Extend dataset (training data)  
Random rotation, inversion and extension of data sets



# Inception Resnet



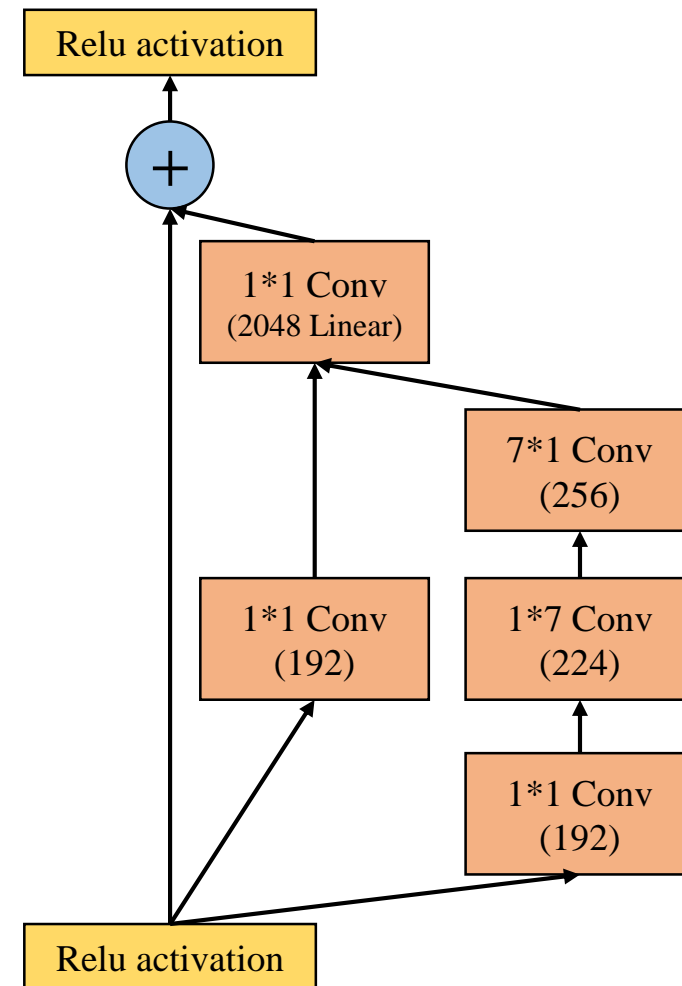
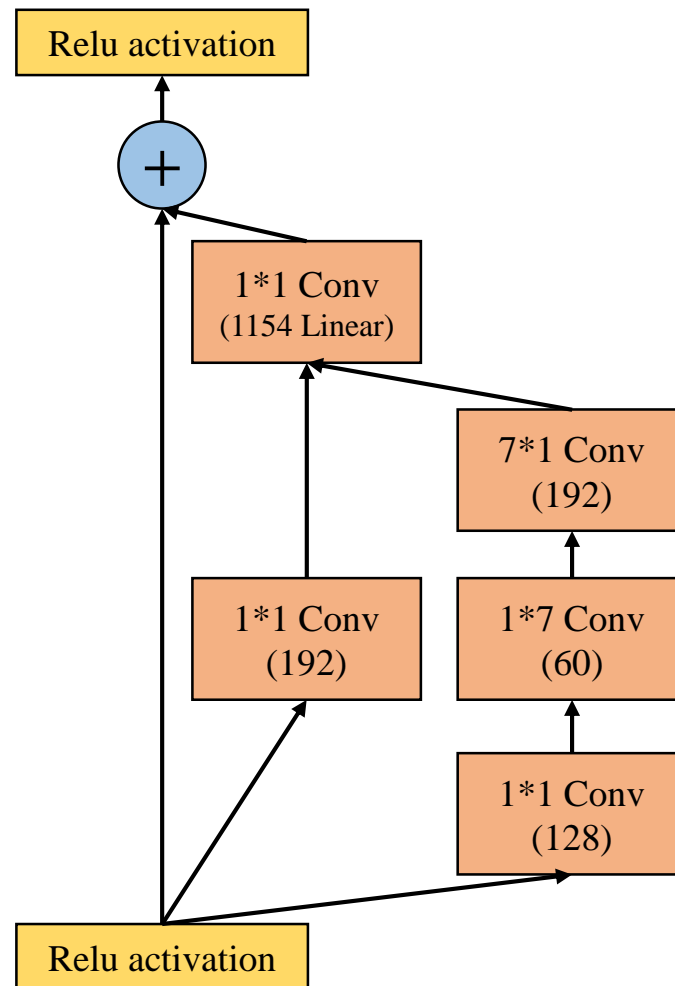
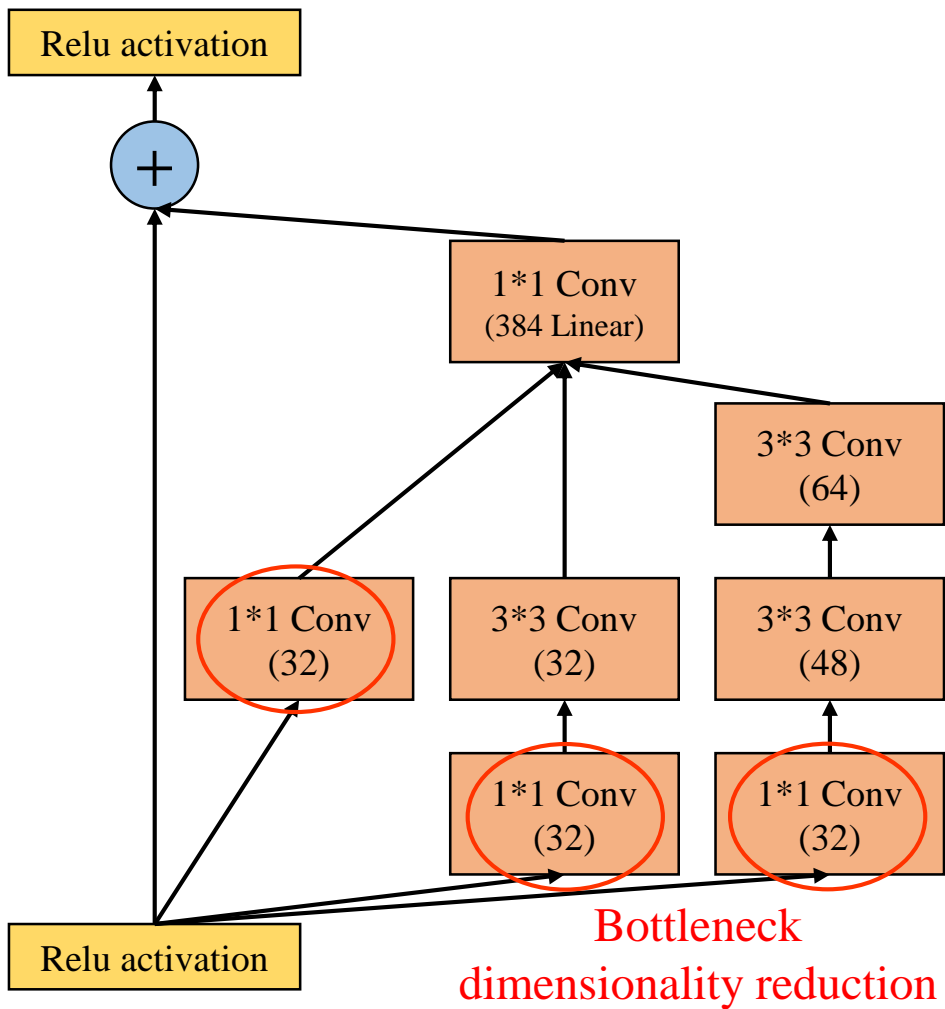
- Residual connections are inherently necessary for training very deep convolutional models.
- However the use of residual connections seems to improve the training speed greatly, which is alone a great argument for their use.



# Inception Resnet

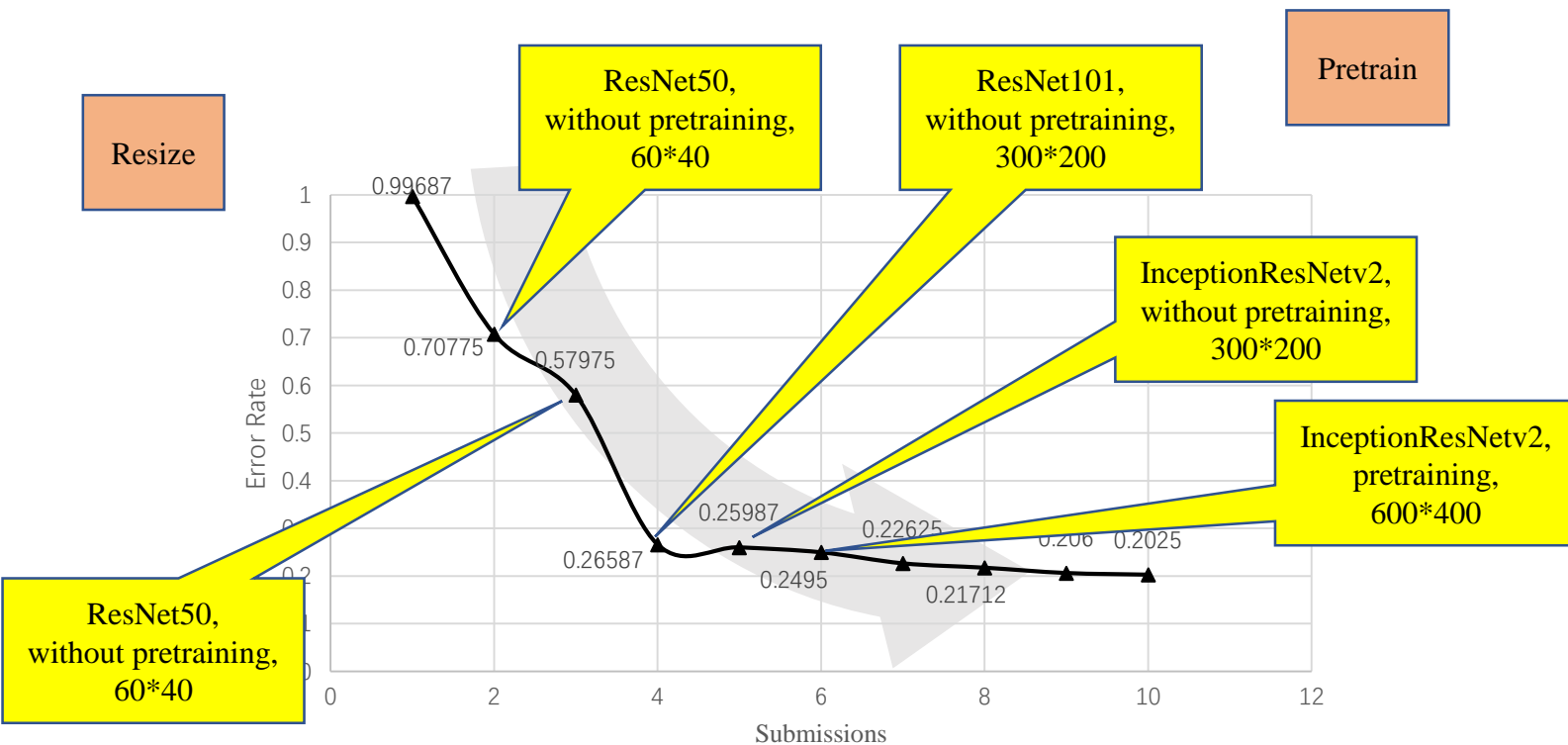


- Inception-Resnet-A/B/C

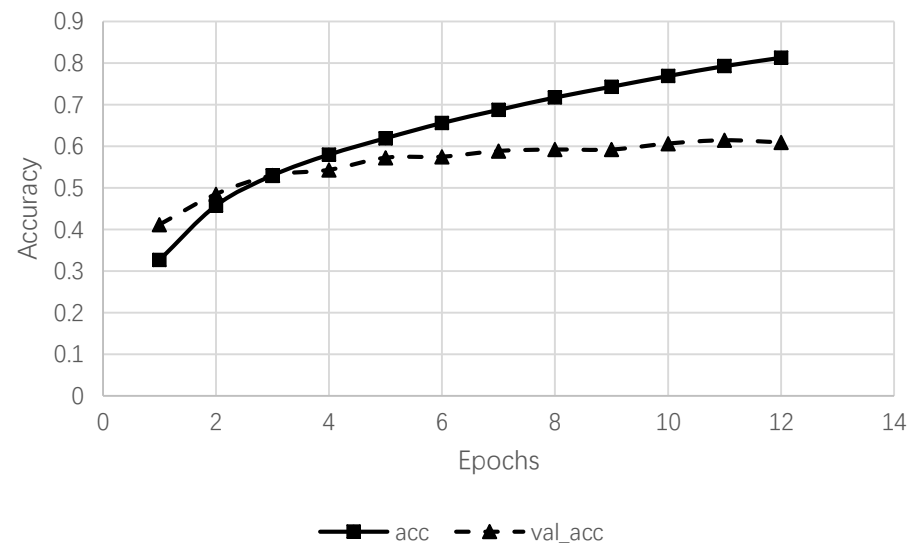
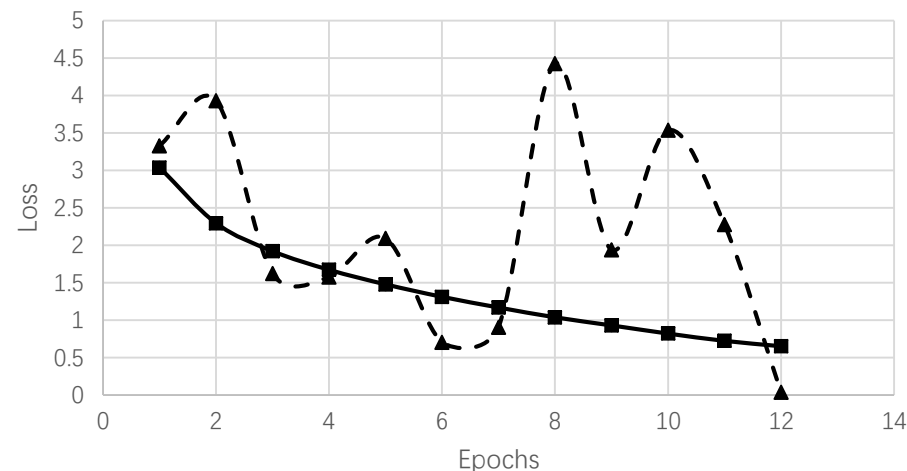


# Evaluation

- Top-1 accuracy based on validation data: 61.42%
- Top-3 accuracy based on testing data: 80.425%



Since we tried a lot of models to fit our dataset, we found that Inception Resnet v2 performed better.





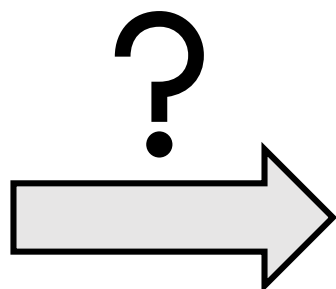
# Conclusion



- Result Analysis:

- With the refinement of pretreatment and the deepening of the network, our model has achieved good results (more than 80%) under the top-3 accuracy.
- But for some reasons, we misclassified some pictures.

test\_0004.jpg



Category 564



Category 725



Category 518



- Future Work:

- Design a model more suitable for this kind of dataset.
- Introduce additional attribute information and knowledge. (e.g. ingredient, brand name)



# Thank you!

Details (code, without dataset) of this project can be found at:  
<https://github.com/chwzuo/ImagesClassification>

Code, data  
pretreatment,  
presentation



Chengwen Zuo  
左成文  
Graduate, M.E.  
Dept. of Electronic Eng.  
Tsinghua University

Code,  
Environment,  
Model



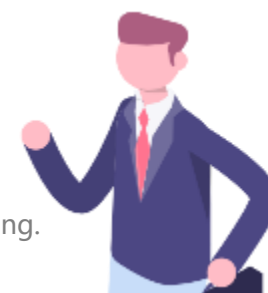
Jin Wang  
王晋  
Graduate, M.E.  
Dept. of Electronic Eng.  
Tsinghua University

Code,  
Evaluation,  
Model



Yuxuan Zheng  
郑禹轩  
Graduate, PhD  
Dept. of Electronic Eng.  
Tsinghua University

Code, Result  
Analysis,  
Report



Chen Cai  
蔡晨  
Undergraduate  
Dept. of Electrical Eng.  
Tsinghua University