李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021)· 课程资料包 @ShowMeAI

视频
中英双语字幕

课件
一键打包下载

笔记
官方笔记翻译

代码
作业项目解析

视频 · B 站 [扫码或点击链接]
https://www.bilibili.com/video/BV1fM4y137M4

课件 & 代码 · 博客 [扫码或点击链接]
http://blog.showmeai.tech/ntu-hylee-ml

Awesome AI Courses Notes Cheatsheets 是 ShowMeAI 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets· 持续更新中			
知识图谱	**图机器学习**	**深度强化学习**	**自动驾驶**
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094

微信公众号

资料下载方式 2：扫码点击底部菜单栏
称为 **AI 内容创作者？** 回复 [添砖加瓦]

Machine Learning HW13

ML TAs
ntu-ml-2021spring-ta@googlegroups.com

Outline

- Task Description
- Dataset
- Guidelines
- Kaggle & Regulations
- Grades
- Links

Links

- Kaggle
- colab tutorial
- video ch / en

Due

- Kaggle: 2021/07/02 23:59:59
- Code & Report: 2021/07/04 23:59:59
- **No Late Submission!!!**

Task Description

- Network Compression: Use a small model to simulate the prediction/accuracy of the large model.
- In this task, you need to train a very small model to complete HW3, that is, do the classification on the food-11 dataset.

Task - Food Classification AGAIN!

- Same as HW3.

Task - Food Classification AGAIN!

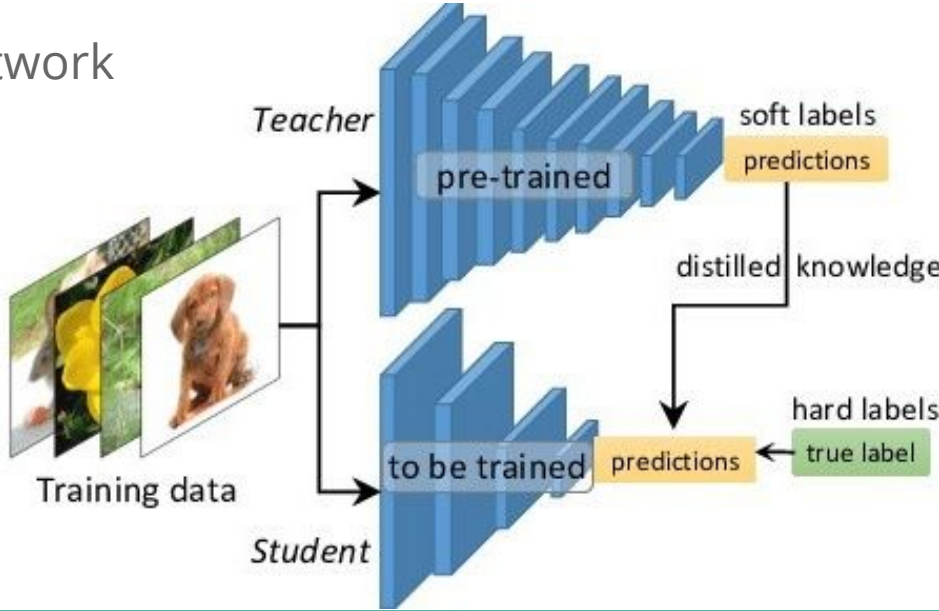
- The images are collected from the food-11 dataset classified into 11 classes.
- The dataset here is slightly modified:
- Training set: 280 * 11 **labeled** images + 6786 **unlabeled** images
- Validation set: 60 * 11 **labeled** images
- Testing set: 3347 images
- DO NOT utilize the original dataset or labels.
    - This is cheating.

Intro

- There are many types of Network/Model Compression, here we introduce two:
  - **Knowledge Distillation**: Let the small model learn better by observing the behavior(prediction) of the large model when learning. (in literal: let the small model extract the knowledge out of the big model)
  - **Design Architecture**: Use fewer parameters to represent the original layer. (E.g. Normal Convolution → Depthwise & Pointwise Convolution)
  - If you are interested in **Network Pruning**, you can view colab tutorial in ML-Spring2020-HW7.

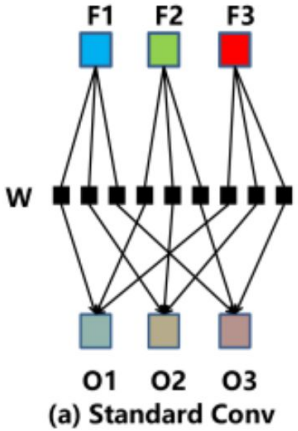
Intro - Knowledge Distillation

- When training a small model, add some information from the large model (such as the probability distribution of the prediction) to help the small model learn better.
- We have provided a well-trained network to help you do knowledge distillation (Acc ~= 0.855).
- Please note that you can only use the pre-trained model we provide when writing homework.

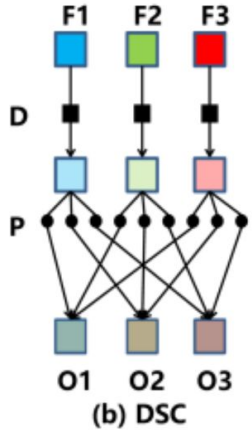


Intro - Design Architecture

- Depthwise & Pointwise Convolution Layer (Proposed in MobileNet)
  - You can consider the original convolution as a Dense/Linear Layer, but each line/each weight is a filter, and the original multiplication becomes a convolution operation. (input*weight → input * filter)
  - Depthwise: let each channel pass a respective filter first, and let every pixel pass the shared-weight Dense/Linear.
  - Pointwise is a 1x1 Conv.
- It is strongly recommended that you use similar techniques to design your model. (NMkk / Nkk+NM)



(a) Standard Conv



(b) DSC

Regulations

- You should NOT plagiarize, if you use any other resource, you should cite it in the reference. (*)
- Do NOT share codes or prediction files with any living creatures.
- Do NOT use any approaches to submit your results more than 5 times a day.
- Do NOT search or use additional data.
- Do NOT search the label or dataset on the Internet.
- **Do NOT use pre-trained models on any image datasets.**
- Your **final grade x 0.9** if you violate any of the above rules.
- Prof. Lee & TAs preserve the rights to change the rules & grades.

Special Regulations - 1

- **Make sure that the total number of parameters of your model should less or equal than 100, 000.**
  - **Please make sure to follow this rule before submitting kaggle / NTU COOL to prevent anyone from polluting the leaderboard.**
  - **If you don't follow this rule, you'll get 0 point in this assignment.**
- **DO NOT USE TEST DATA FOR PURPOSE OTHER THAN INFERENCING.**
  - Because If you use teacher network to predict pseudo-labels of the test data, you can only use student network to overfit these pseudo-labels without train/unlabeled data.
    In this way, your kaggle accuracy will be as high as the teacher network, but the fact is that you just overfit the test data and your true testing accuracy is very low.
  - These contradict the purpose of these assignment (network compression); therefore, you should not misuse the test data.
  - If you have any concerns, you can email us.

Special Regulations - 2

- We strongly recommend that you use the torchsummary package to measure the number of parameters of your model. Note that non-trainable parameters should also be considered.

- **Ensemble techniques / (or any other multi-model techniques) are allowed.** But you need to sum all numbers of the parameters and make sure this number is not exceed 100,000.

Grades

- +3pt : code submission
- +1pt : NTU cool questions (Describe your model & it's # of parameters)
- +1pt : Simple public baseline
- +1pt : Simple private baseline

- +1 : Medium public baseline
- +1 : Medium private baseline

- +0.75 : Strong public baseline
- +0.75 : Strong private baseline

- +0.25 : Boss public baseline
- +0.25 : Boss private baseline

Baseline Guides

- Simple Basline (2pts, acc ≥ 0.59856, 2 hour)
  - Just run the code and submit answer.

- Medium Baseline (2 pts, acc ≥ 0.65412, 2 hours)
  - Complete the loss in knowledge distillation and control alpha & T.

- Strong Baseline (1.5 pts, acc ≥ 0.72819, 4 hours)
  - Modify model architecture with depth- and point-wise convolution layer.
    - Or, you can take great ideas from MobileNet, ShuffleNet, DenseNet, SqueezeNet, GhostNet, etc.
  - Any techniques and methods you learned in HW3 - CNN. For example, make data augmentation stronger, modify semi-supervised learning, etc.

Baseline Guides

- Boss Baseline (0.5 pts, acc ≥ 0.81003)

    - Make your teacher net more stronger.
        - If your teacher net is too strong, you can consider TAKD techniques.
    - Implement other advanced knowledge distillation.
        - For example, DML, Relational KD....
    - If the number of the parameters of your model is slightly larger than the constraint (100, 000), you can use network pruning.
    - If you got confused of previous techniques, you can check out TA's lesson in last year. (slides, video)

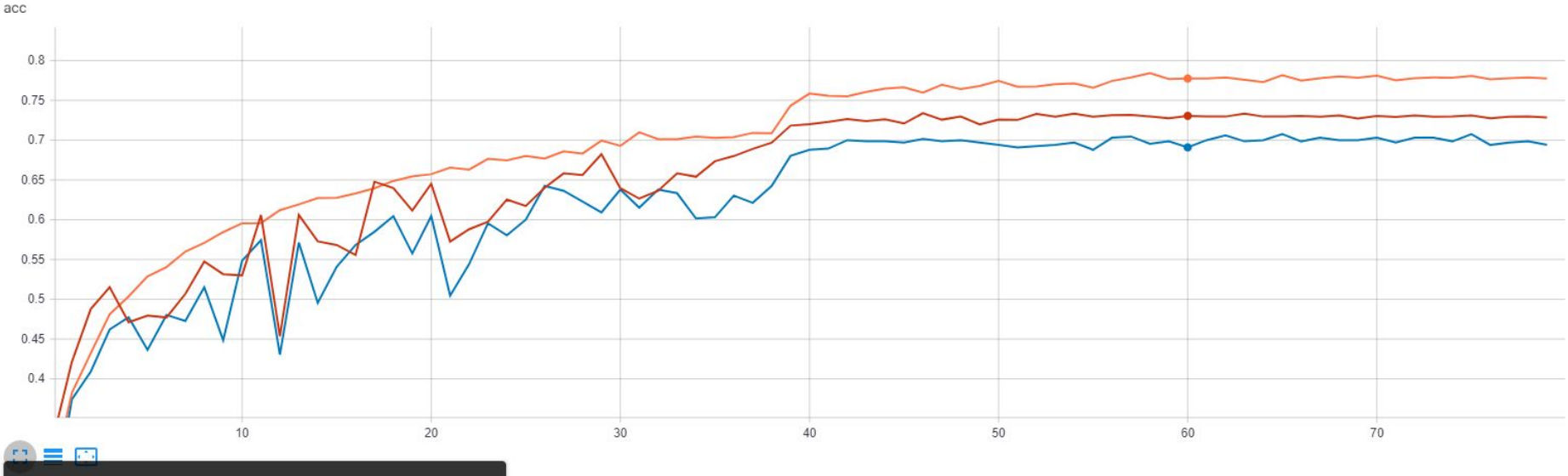
Code Submission (3pts)

- Submit on NTU COOL
  - Deadline: 7/4 (Sun.) 23:59
  - Compress your code and report into <student_ID>_hw13.zip（e.g. b10123456_hw13.zip）
  - We can **only** see your **last submission**.
  - DO NOT submit your model or dataset.
  - If your code is not reasonable, your semester grade x 0.9.
- Your .zip file should include only
  - Code: either .py or .ipynb
  - Report: .pdf (only for those who got 10 points)
- If you successfully get 10 points in this assignment,
  - You can submit a report in PDF format briefly describing what you have done for extra 0.5 pts.
  - Report template

Question Submission (1pts)

- Submit on NTU COOL
  - Deadline: 7/4 (Sun.) 23:59
  - We can **only** see your **last submission**.
  - For more information about this question, see NTU COOL.

Learning Curve



Name	Smoothed	Value	Step	Time	Relative
acc\test	0.7305	0.7305	60	Fri Jun 11, 05:02:26	1h 30m 59s
acc\train	0.7774	0.7774	60	Fri Jun 11, 05:02:26	1h 30m 59s
acc\valid	0.6909	0.6909	60	Fri Jun 11, 05:02:26	1h 30m 59s

If any questions, you can ask us via...

- NTU COOL (recommended)
  - [Link]
- Email
  - ntu-ml-2021spring-ta@googlegroups.com
  - The title should begin with “[hw13]”
- TA hour
  - Each Monday 19:00~21:00 @Google Meets
  - Each Friday 13:30~14:20 Before Class @Google Meets

李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021)· 课程资料包 @ShowMeAI

视频
中英双语字幕

课件
一键打包下载

笔记
官方笔记翻译

代码
作业项目解析

视频 · B 站 [扫码或点击链接]
https://www.bilibili.com/video/BV1fM4y137M4

课件 & 代码 · 博客 [扫码或点击链接]
http://blog.showmeai.tech/ntu-hylee-ml

Awesome AI Courses Notes Cheatsheets 是 ShowMeAI 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets· 持续更新中			
知识图谱	**图机器学习**	**深度强化学习**	**自动驾驶**
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094

微信公众号

资料下载方式 2：扫码点击底部菜单栏
称为 **AI 内容创作者？** 回复 [添砖加瓦]