인공지능 개론
L11 Vision Applications

국민대학교
소프트웨어융합대학원
박하명

컴퓨터 비전의 대표적 문제들

1. Classification
2. Object Detection
3. Semantic Segmentation
4. Pose Estimation

Classification

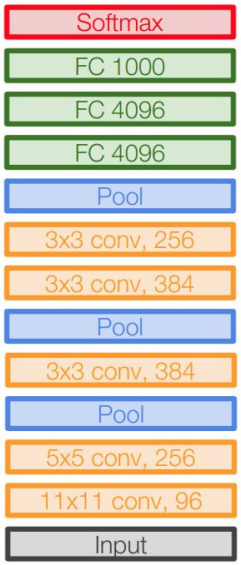
- 이미지가 주어지면, 그게 무슨 이미지인지 맞추기

뭘까요? **토끼!**

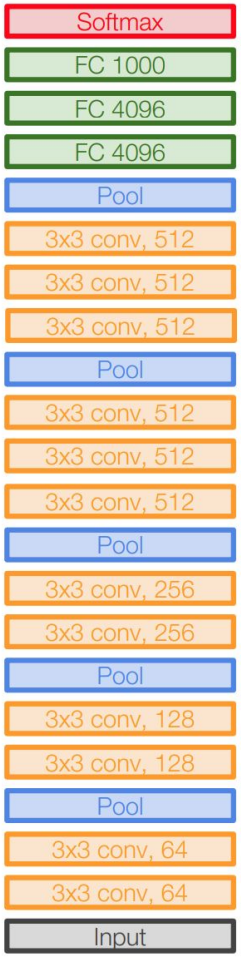
뭘까요? **고양이**

뭘까요? **강아지**

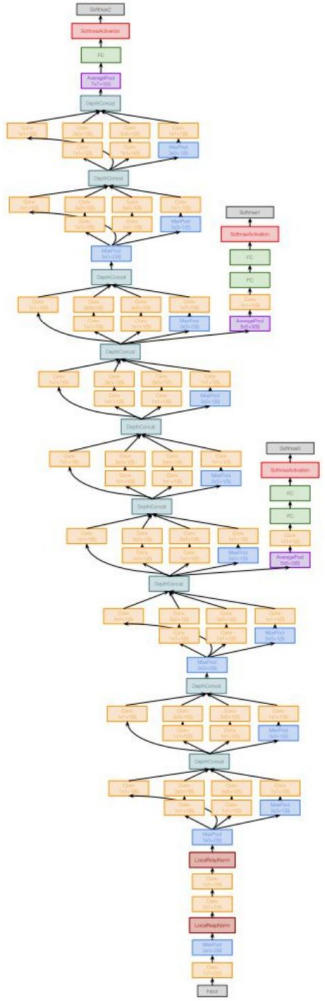
Models for Classification



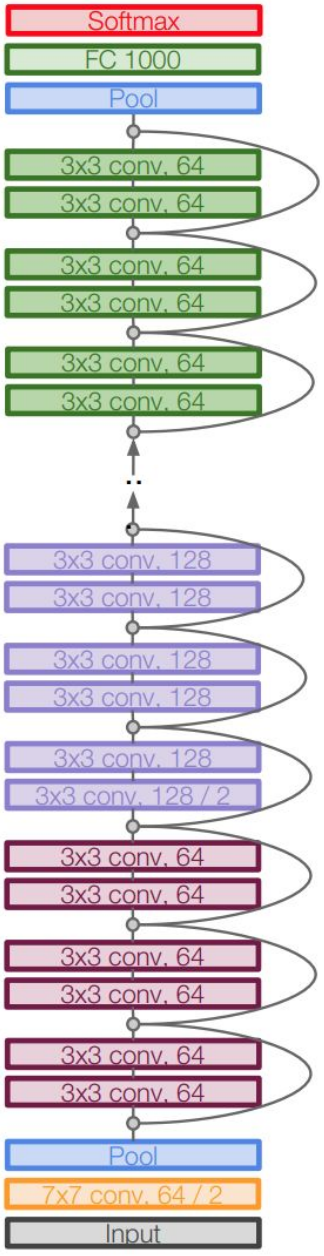
AlexNet



VGG



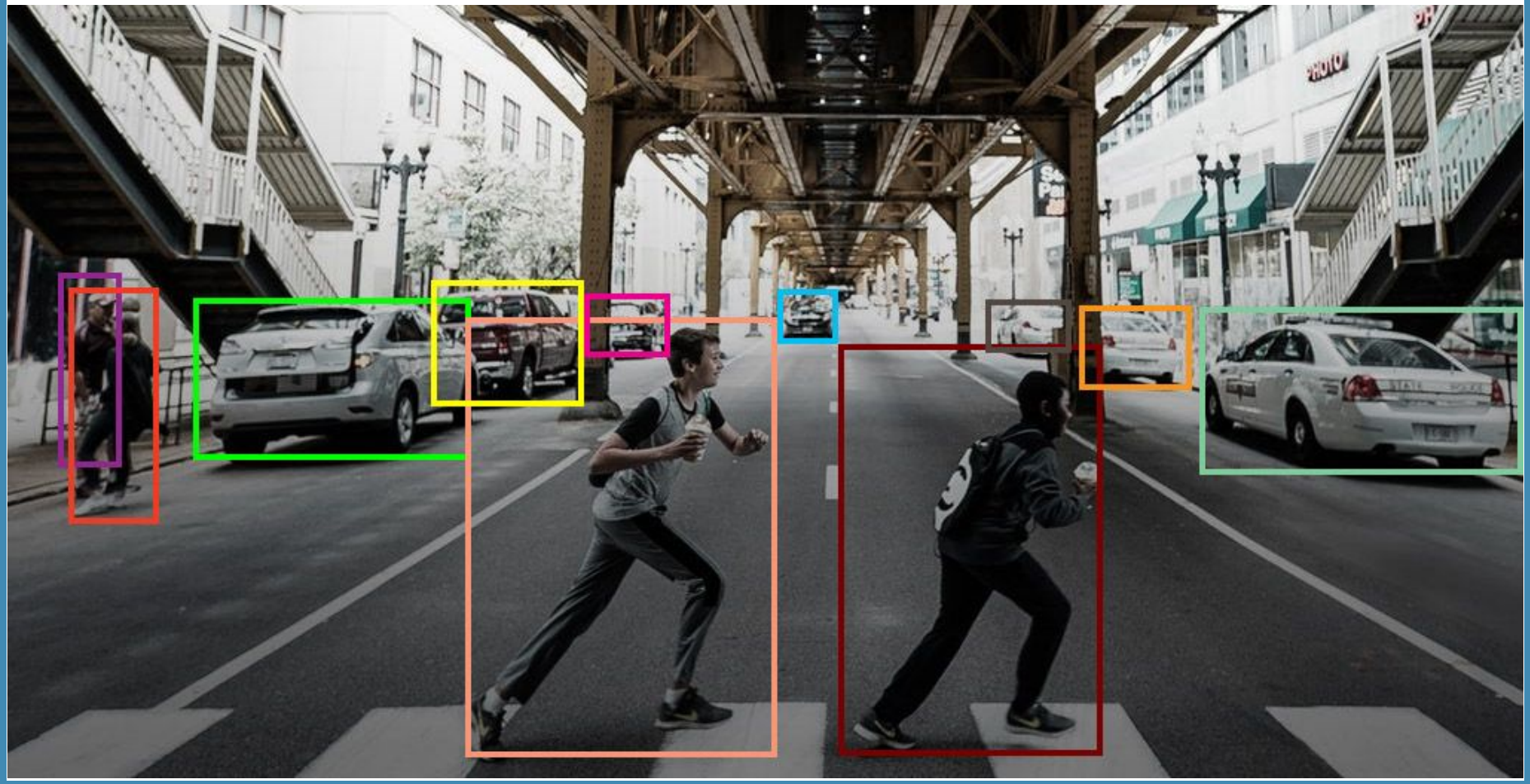
GoogLeNet



ResNet

Object Detection

- 이미지에서 **개체들을 찾아내기** (무엇이 어디에?)



Models for Object Detection

- object가 있는 영역을 먼저 찾음
- 각 영역마다 classification 수행

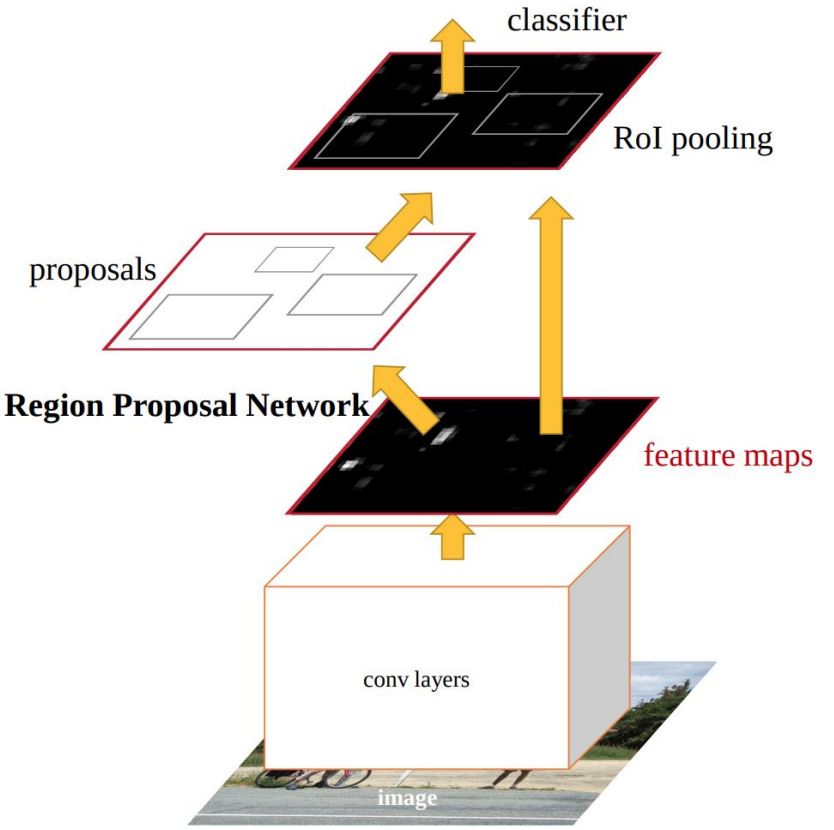
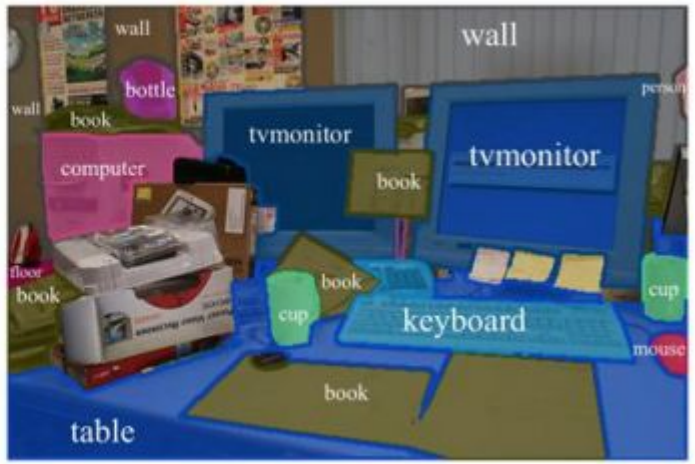


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

Semantic Segmentation

- 이미지를 의미(semantic) 단위로 여러 영역으로 나눔

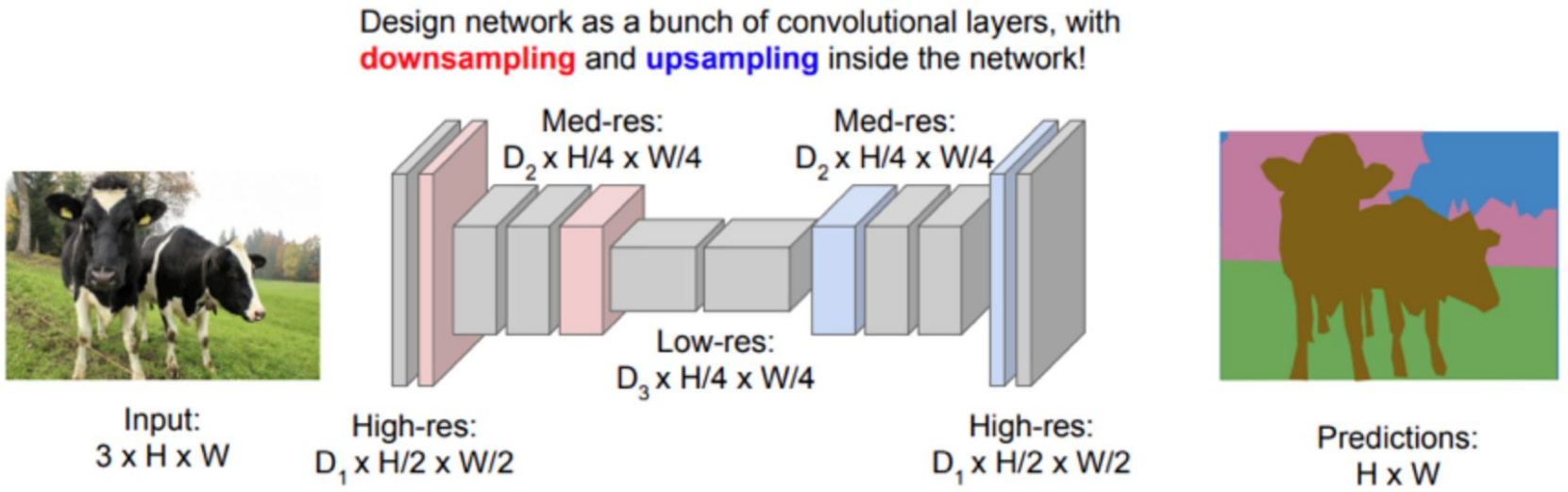
Semantic Segmentation 응용

- 구글 Pixel 2 스마트폰의 인물모드 촬영

https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html

Models for Semantic Segmentation

- 픽셀 단위로 classification



segmented

1: Person
2: Purse
3: Plants/Grass
4: Sidewalk
5: Building/Structures

3 3 3 3 3 3 3 3 3 3 3 3 5 5 5 5 5 5
3 3 3 3 3 3 3 3 3 3 3 3 5 5 5 5 5 5
3 3 3 3 3 3 1 1 3 3 3 3 5 5 5 5 5 5
3 3 3 3 3 1 1 1 1 3 3 3 5 5 5 5 5 5
3 3 3 3 3 3 1 1 3 3 3 5 5 5 5 5 5 5
5 5 3 3 3 3 1 1 3 3 5 5 5 5 5 5 5 5
4 4 3 4 1 1 1 1 1 1 4 4 4 5 5 5 5 5
4 4 3 4 1 1 1 1 1 1 4 4 4 4 4 5 5 5
4 4 4 1 1 1 1 1 1 1 1 4 4 4 4 4 4 4
3 3 3 1 1 1 1 1 1 1 1 4 4 4 4 4 4 4
3 3 3 1 2 2 1 1 1 1 1 4 4 4 4 4 4 4
3 3 3 1 2 2 1 1 1 1 1 4 4 4 4 4 4 4

Input

Semantic Labels

Person Keypoint Detection

- Pose Estimation 이라고도 함
- 사람의 눈, 코, 귀, 팔, 다리 등의 위치를 찾아내기

Models for PKD

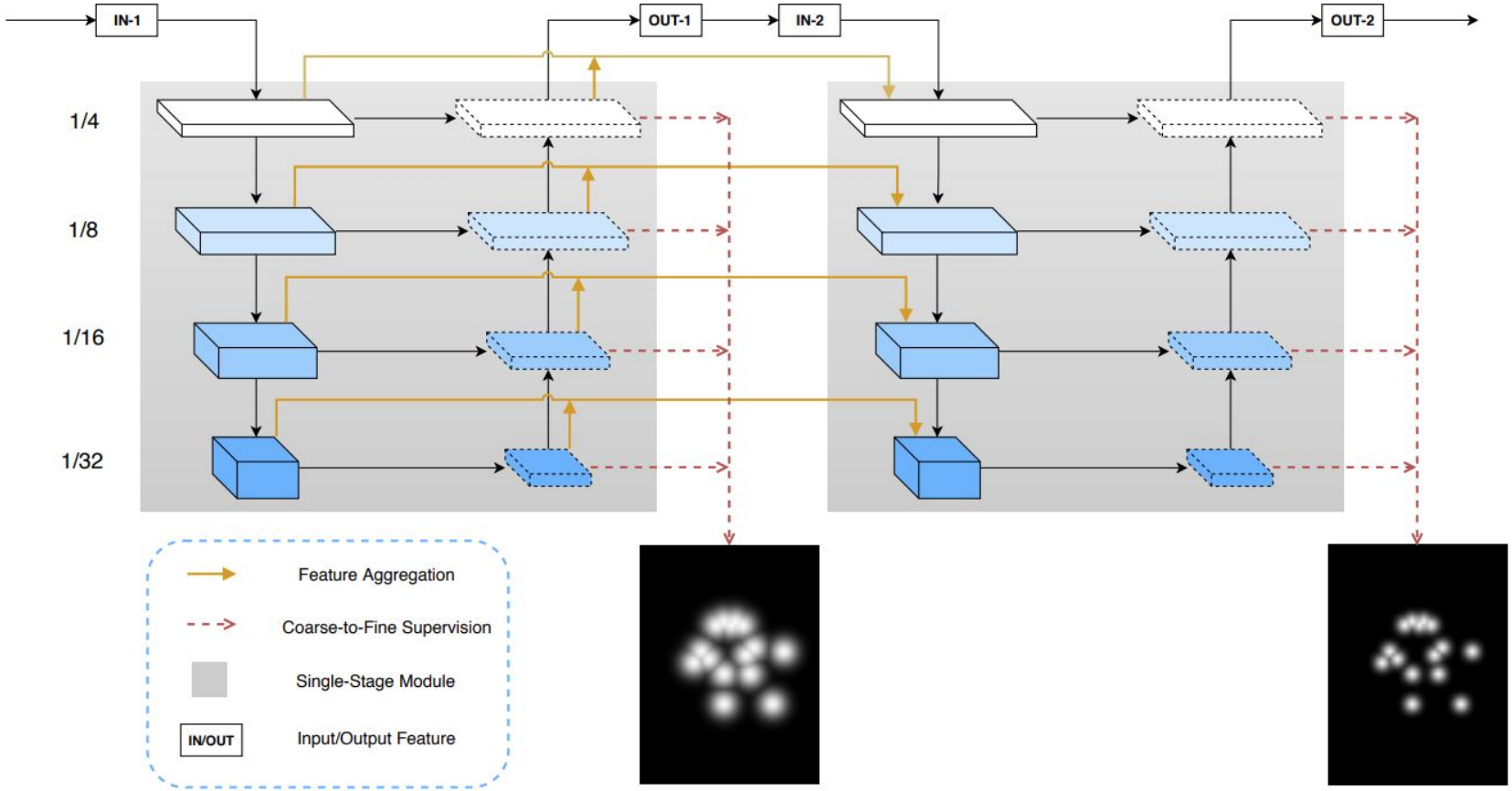
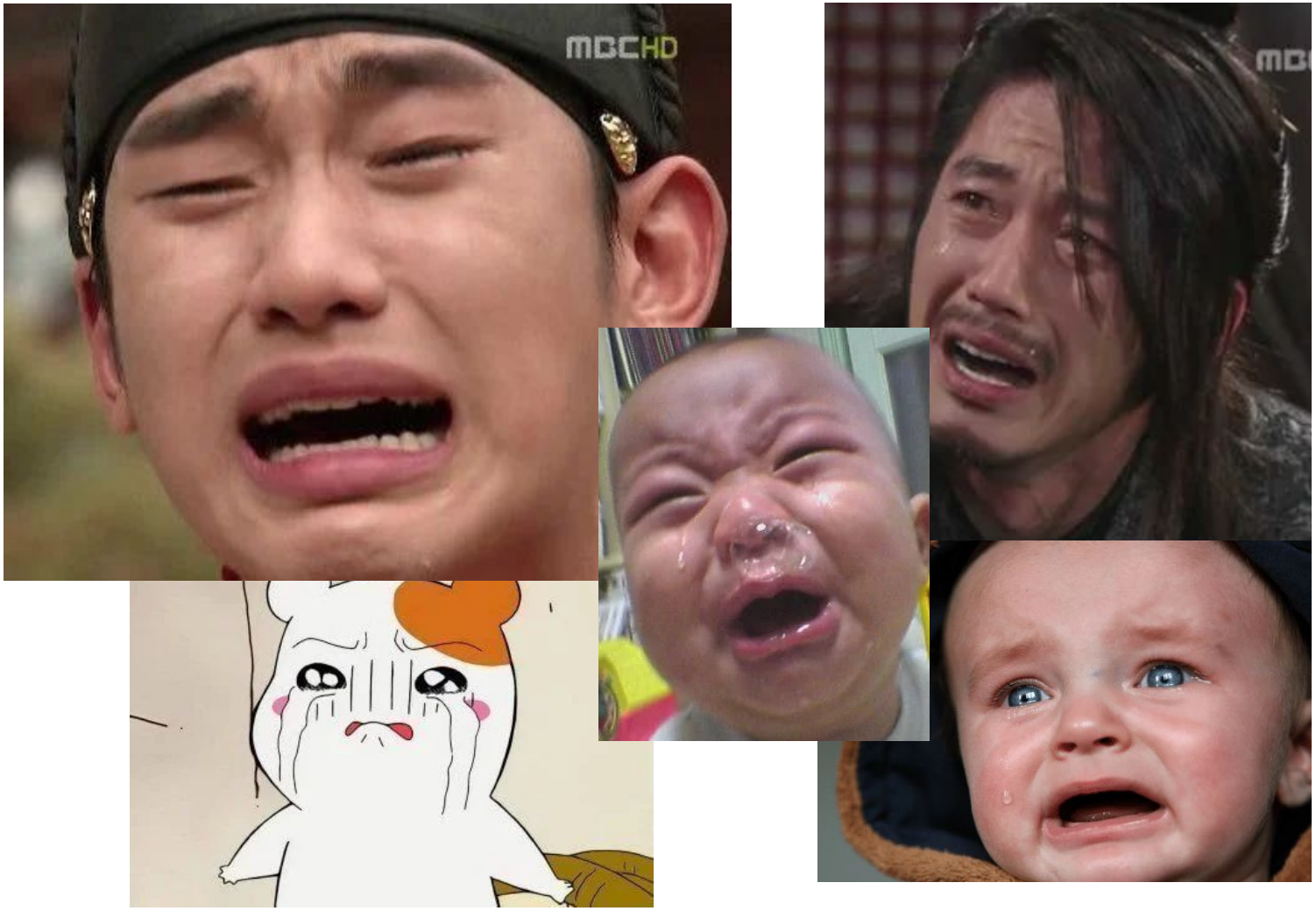


Figure 2. Overview of Multi-Stage Pose Network (MSPN). It is composed of two single-stage modules. A cross stage aggregation strategy (zoomed in Figure 3) is adopted between adjacent stages (Section 3.2). A coarse-to-fine supervision strategy further improves localization accuracy (Section 3.3).

https://arxiv.org/pdf/1901.00148v4.pdf

이걸 다 어떻게 해...ㅠㅠ



Torchvision

- 컴퓨터 비전을 위한 인기있는 **데이터셋**, **모델**, **이미지 변환 도구** 등을 제공
- **torchvision.datasets**: MNIST, CIFAR, ImageNet, ...
- **torchvision.io.Video**: 영상 입출력 도구
- **torchvision.models**: 유명한 모델의 구조 및 pretrained된 모델 제공
  - **Classification**: AlexNet, VGG, ResNet, GoogLeNet …
  - **Semantic Segmentation**: FCN, DeepLabV3
  - **Object Detection, Person Keypoint Detection**: FPN
  - **Video Classification**: ResNet 3D
- **torchvision.transforms**: 이미지 변환 도구

Question?