CS2205.CH1501
PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC

GVHD: PGS. TS Lê Đình Duy

Nhóm:

Mai Phương Nga - CH2001010

Nguyễn Như Thanh - CH2001015

Trần Hiếu Đại - CH2001001

Nội dung

1. Giới thiệu paper
2. Bài toán
3. Vấn đề của bài toán
4. Ý tưởng và phương pháp giải quyết
5. Thách thức

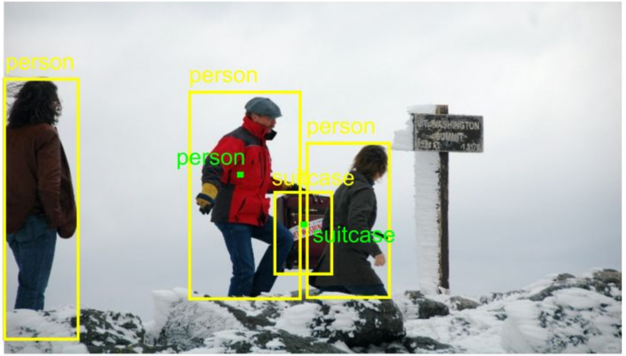
1. Giới thiệu paper

- Paper: "**Fast Object Class Labelling via Speech**", ***Michael Gygli, Vittorio Ferrari;*** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 5365-5373
- Link: https://openaccess.thecvf.com/content_CVPR_2019/papers/Gygli_Fast_Object_Class_Labelling_via_Speech_CVPR_2019_paper.pdf
- Google scholar: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Fast+Object+Class+Labelling+via+Speech&btnG=

2. Bài toán

- “Object class labelling is the task of **annotating images with labels** on the presence or absence of objects **from a given class vocabulary.**”
- “Ask a separate yes/no question for each class of a given vocabulary.”
- “Deep neural networks **need millions of training examples to obtain high performance.**”

→ Need to reduce the time of object labelling tasks.



Object class labels: person, suitcase

3. Vấn đề bài toán

- Cost (time consuming) of labelling object classes task on large datasets is high.
  - “scales linearly in the size of the vocabulary” → “becomes **very inefficient when the vocabulary is large**”.

    E.g: ILSVRC dataset: “getting labels for the **200 object classes** in the vocabulary would take close to **6 minutes per image despite each image containing only 1.6 classes on average**”

4. Ý tưởng & phương pháp giải quyết

- Idea: labels object class **by using speech**.
    - “we propose a new interface where classes are annotated via speech”
- Methods:
    - “Given an image, annotators scan it for objects and mark one per class by **clicking on it and saying its name**”
    - “combining speaking with pointing”: “when using multimodal interfaces, people naturally choose to point for providing spatial information and to speak for semantic information”

4. Ý tưởng & phương pháp giải quyết (tt)

- Methods (cont.):
  - “obtain object class labels by associating audio segments to clicks and then transcribing the audio“
  - “we rely on Google’s automatic speech recognition API“

5. Thách Thức

In order to reliably transcribe speech to text, several technical challenges need to be tackled, such as **segmenting the speech and obtaining high-accuracy transcriptions**. Furthermore, as speech is free-form in nature, **annotators need to be trained** to know the class vocabulary to be annotated in order to not label other objects or forget to annotate some classes.