

懒人制作学术会议Oral/Spotlight Video指南

山佳木又 CVer 5月25日

点击上方“CVer”，选择加“星标”或“置顶”
重磅干货，第一时间送达

本文作者：山佳木又

<https://zhuanlan.zhihu.com/p/142394787>

本文已由原作者授权，不得擅自二次转载

引言

在疫情影响下，不少学术会议都变成了线上举行，于是乎制作在线上会议上使用的oral视频成了科研工作者们的新任务，最近做了BBN工作CVPR2020 oral材料，slides的制作比较简单，有很多帖子可以参考，写个文章记录下在mac OS下做视频的工具和思路（硬广）。

学术会议的视频中，图像一般是slides，声音一般是对slides的讲解。（虽然和在现场分享别无二致，但是没有实体听众，多多少少会缺点人情味和紧张感~）一个很自然的思路是，自己线下配合slides，在小黑屋里边做presentation，边录屏、录音。

虽然这样看起来流程非常自然，但是实操过程中会经常失败，比如时不时slides动画的切换没有和嘴巴配合好，时不时嘴巴秃噜了念了个错误的词，抑或是对自己的pronunciation不是很自信。如此，想录一个完整的pre出来是时间成本较高的一件事。因为我是个比较懒的人，所以想了如下的懒人思路：

1. 写好一份精炼的讲稿，由于正常情况下人一分钟能说130~150词，所以讲稿的长度完全由视频要求的时长决定，由于CVPR视频限时5分钟，我写的就是700词左右的讲稿；
2. 结合讲稿做好静态的slides，在脑袋里模拟一下有哪些地方需要用动画配合讲解，再添加上动画，用latex做slides的大神除外；
3. 把讲稿扔进text-to-speech软件里，生成一份由AI念的稿子，录下来存成mp3格式；
4. 配合AI念的语音，完成对slides的录屏，存成mp4格式；
5. 把语音和录屏剪辑在一起，完成啦！

为什么要这么做呢？原因有以下几点：

1. 先写稿子，稿子决定了pre的质量，稿子可以反复修改，操作空间巨大；
2. 写好稿子再做slides，速度会非常快，而且思路会更清晰；
3. AI生成的念稿语音可以解放我们的嘴巴，专心做好slides的页面切换和动画配合；

4. 完成录屏后，可以直接用AI生成的语音合成视频，也可以自己跟读AI的语音，同时录音。
跟读要容易很多很多，实操一把就知道；

剩下的文章分步骤详细讲讲怎么做。

讲稿

选择自己喜欢的写作方式，可以写俏皮一点也可以正规一点，感觉圈子还是很包容的！这里放一小段，可以用grammarly之类的app改一改。

CVPR pre draft

Opening

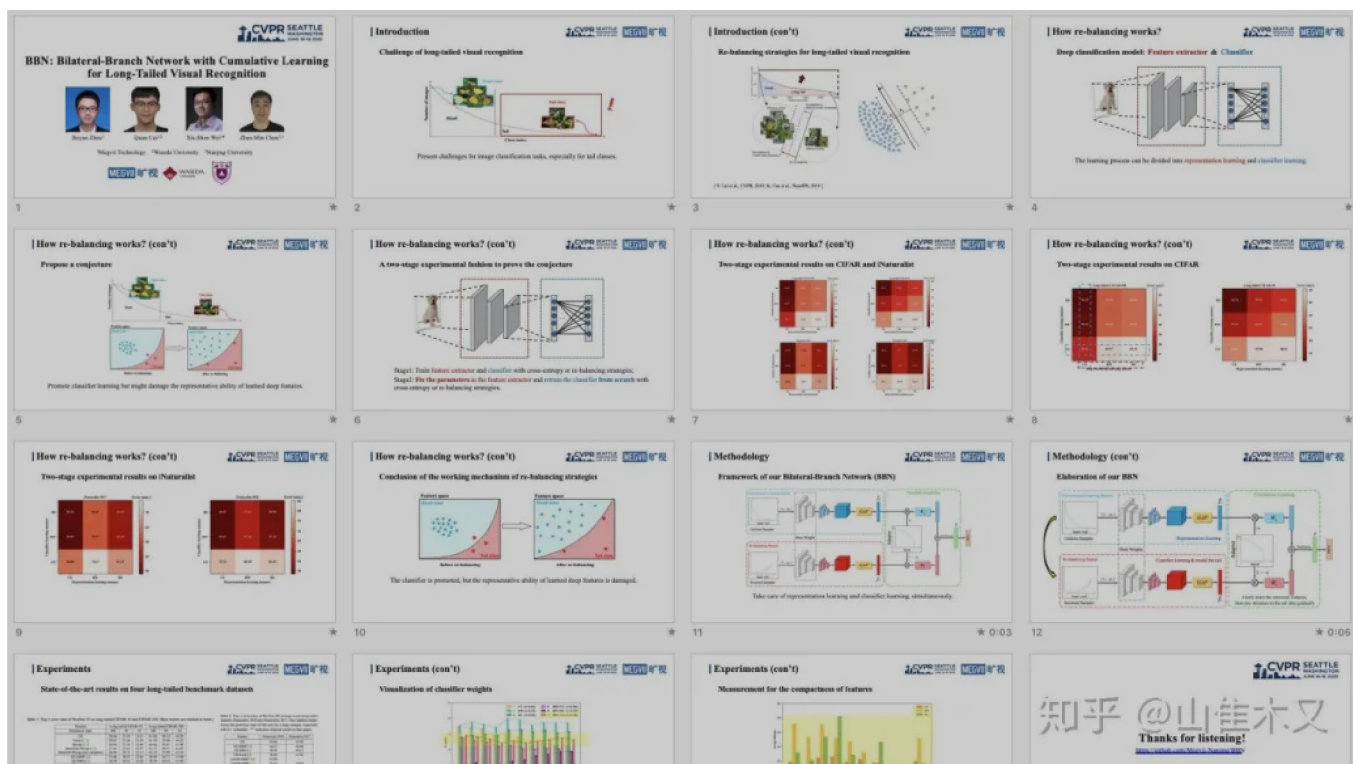
Hello everyone thanks for watching this video. Today I'm going to talk about our work BBN, which is a bilateral-branch network for long-tailed visual recognition.

Long-tailed distributions could be a natural property of real-world large-scale datasets. The extreme imbalance can present tremendous challenges for the image classification task, especially for the tail classes. Re-balancing strategies are proposed to solve the long-tailed visual recognition task by re-sampling the examples or reweighting the losses of examples within mini-batches. So that, the data distribution of training and test sets could become close.

知乎 @山雀木文

Slides

知乎上有很多帖子教怎么做「学术ppt」，搜搜就有啦！主要风格还是简洁，不要摆太多字在slides上就好，TL;DR ~



AI念稿（语音）

这里推荐谷歌家的text-to-speech，谷歌牛逼我只能说，太逼真啦！试用功能就够用啦。下面图中的红框可以调节语速，这个功能可以让你把稿子的时长刚好控制在5分钟，也是非常节省时间的一步，不需要自己瞎琢磨语速。

立即将文字转换为语音

输入所需内容，选择一种语言，然后点击“Speak It”即可收听。

Text to speak:

Google Cloud Text-to-Speech enables developers to synthesize natural-sounding speech with 100+ voices, available in multiple languages and variants. It applies DeepMind's groundbreaking research in WaveNet and Google's powerful neural networks to deliver the highest fidelity possible. As an easy-to-use API, you can create lifelike interactions with your users, across many applications and devices.

text ssm1

Language / locale

English (United States)

Voice type

WaveNet

Voice name

en-US-Wavenet-D

Audio device profile

Default

Speed:

1.00

Pitch:

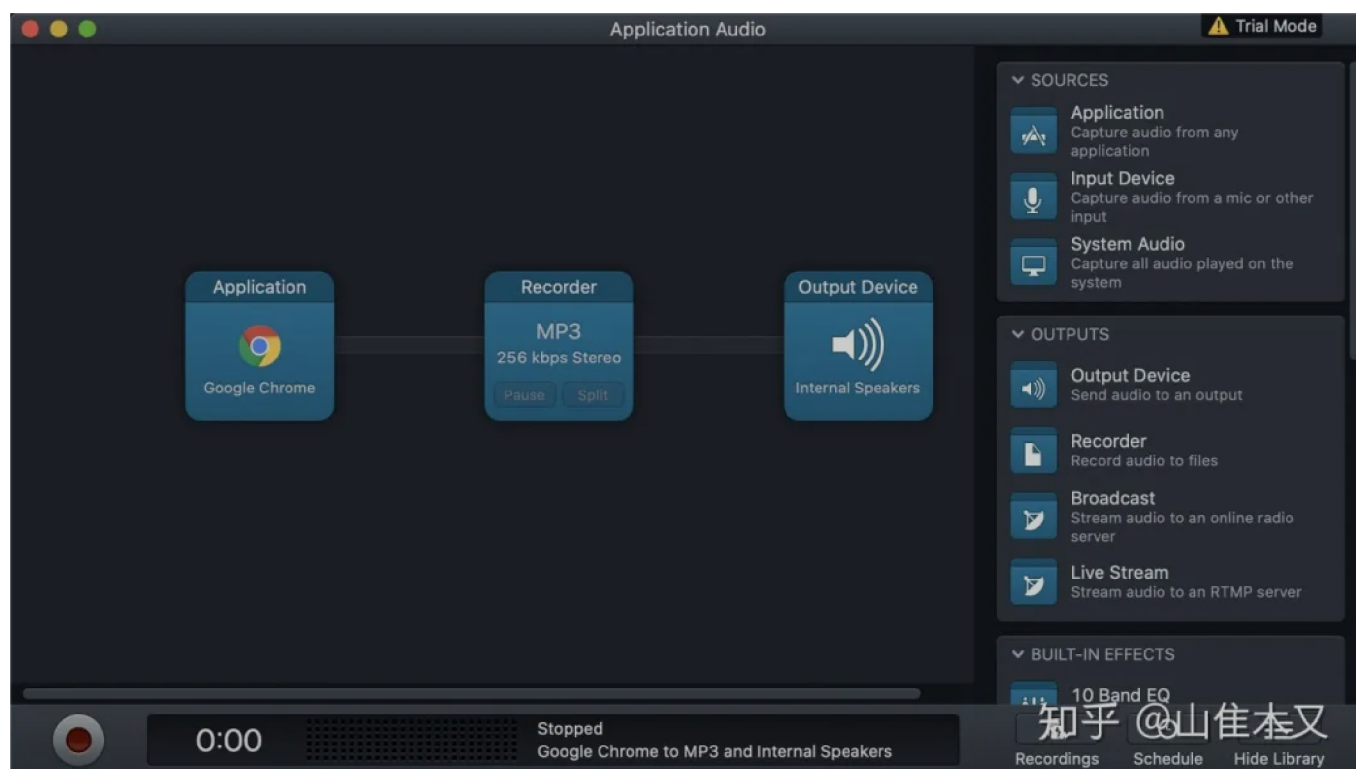
0.00

Show JSON

► SPEAK IT

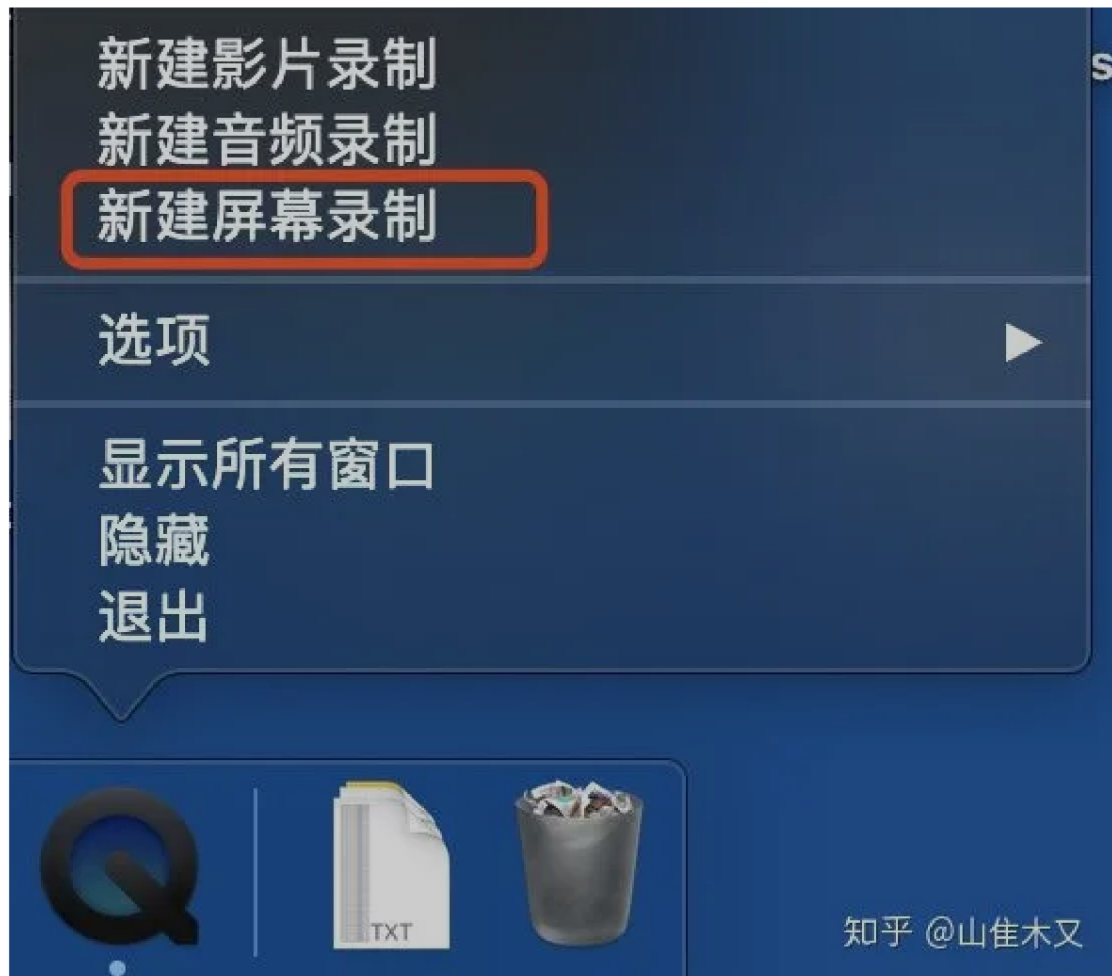
知乎 @山佳木又

由于mac的录屏没有声音，所以这一步会用到一个工具叫作Audio Hijack，这个软件长成下面的样子，可以捕捉app的声音，选择成捕捉浏览器的声音，就可以把谷歌AI念的语音导出成mp3文件咯。



slides录屏（图像）

这里试用mac OS自带的QuickTimePlayer就ok，简单易用，导出的视频是mp4格式的，完美；



剪辑（语音+图像->出货）

强烈推荐mac OS自带的iMovie，几乎没有学习成本，把录屏的mp4和录音的mp3导入这个app，裁剪一下超时的视频和音频，对齐一下时间轴，随后就可以导出成成品视频啦！放一小段看看效果吧！

00:36

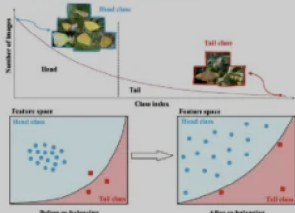
当然，各位也可以把语音换成自己跟读AI的录音，会更有人情味一点。

硬广

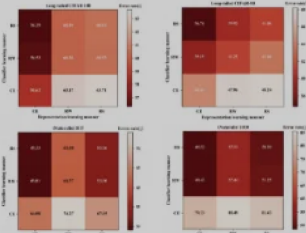
做了poster但是cvpr官方通知不需要了，觉得有点浪费！放在这里宣传一下我们的工作！code已开源！

<https://github.com/Megvii-Nanjing/BBN>

Introduction & Observations



Our work focuses on tackling the visual recognition task of long-tailed data distribution. In this paper, we reveal the mechanism of re-balancing strategies to significantly promote classifier learning but unexpectedly damage the representative ability of the deep features.



In this figure, "CE" (Cross-Entropy), "RW" (Re-Weighting) and "RS" (Re-Sampling) are the conducted learning manners.

- ✓ When fixing the representation (comparing error rates of three blocks in the vertical direction), error rates of classifiers trained with RW/RS are reasonably lower than CE.
- ✓ When fixing the classifier (comparing error rates in the horizontal direction), the representations trained with CE surprisingly get lower error rates than those with RW/RS.

Proposed Bilateral-Branch Network & Cumulative Learning Strategy

Our BBN consists of three main components, i.e., **conventional learning branch**, **re-balancing branch** and **cumulative learning strategy**. Both branches use the same residual network structure and share all the weights except for the last residual block.

✓ Conventional Learning Branch:

A uniform sampler is applied to obtain sample (x_c, y_c) as the input data from the original data distribution.

✓ Re-Balancing Branch:

A reversed sampler is employed to acquire sample (x_r, y_r) according to the following manner:

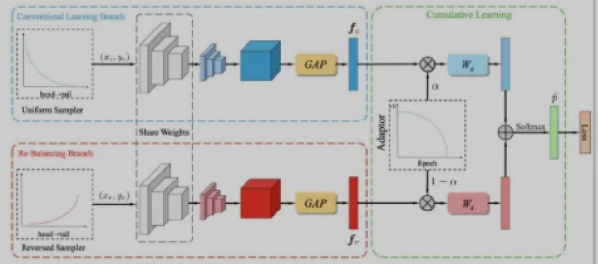
$$P_i = \frac{w_i}{\sum_{j=1}^C w_j}$$

where P_i denotes the sampling possibility for the i -th class, $w_i = \frac{1}{N_i}$ and C is the number of classes

✓ Cumulative Learning Strategy:

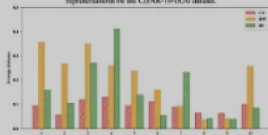
It is designed to first learn the universal patterns and then pay attention to the tail data gradually by controlling the weights for features produced by two branches and the classification loss:

$$\alpha = 1 - \left(\frac{T}{T_{max}} \right)^2 \quad z = \alpha W_c^T f_c + (1 - \alpha) W_r^T f_r$$
$$\mathcal{L} = \alpha E(p, y_c) + (1 - \alpha) E(p, y_r)$$



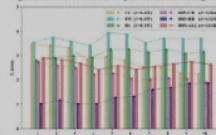
Experiments & Visualization Results

Figure 1. Histogram of the measurement for the comparison of intra-class representations on the CIFAR-10/100 dataset.



Dataset	Class	CE	RW	RS	BBN
CIFAR-10	Head	19.62	15.38	15.41	15.38
	Tail	29.62	13.38	13.54	13.54
	Mean	24.62	14.38	14.48	14.46
	Standard Mean	27.04	12.85	12.97	12.95
CIFAR-100	Head	31.05	15.41	15.41	15.41
	Tail	25.47	10.19	10.19	10.19
	Mean	28.27	12.85	12.85	12.85
	Standard Mean	30.18	11.84	11.84	11.84

Figure 2. L2-norm of classifier weights for different learning manners.



Dataset	Class	CE	RW	RS	BBN
CIFAR-100	Head	19.62	15.38	15.41	15.38
	Tail	29.62	13.38	13.54	13.54
	Mean	24.62	14.38	14.48	14.46
	Standard Mean	27.04	12.85	12.97	12.95
CIFAR-1000	Head	31.05	15.41	15.41	15.41
	Tail	25.47	10.19	10.19	10.19
	Mean	28.27	12.85	12.85	12.85
	Standard Mean	30.18	11.84	11.84	11.84

Contributions & Conclusions

- ✓ For studying long-tailed problems, we explored how class re-balancing strategies influenced representation learning and classifier learning of deep networks, and revealed that they can promote classifier learning significantly but also damage representation learning to some extent.
- ✓ Motivated by this, we proposed a Bilateral-Branch Network (BBN) with a specific cumulative learning strategy to take care of both representation learning and classifier learning for exhaustively improving the recognition performance of long-tailed tasks.
- ✓ By conducting extensive experiments, we proved that our BBN could outperform state-of-the-art results on long-tailed benchmarks, including the large-scale datasets iNaturalist17 and iNaturalist18.
- ✓ In the future, we attempt to tackle the long-tailed recognition problem with our BBN model.

重磅！CVer-论文写作与投稿 交流群已成立

扫码添加CVer助手，可申请加入CVer-论文写作与投稿 微信交流群，目前已满1900+人，旨在交流顶会（CVPR/ICCV/ECCV/ICML/ICLR/AAAI等）、顶刊（IJCV/TPAMI等）、SCI、EI等写作与投稿事宜。

同时也可申请加入CVer大群和细分方向技术群，细分方向已涵盖：目标检测、图像分割、目标跟踪、人脸检测&识别、OCR、姿态估计、超分辨率、SLAM、医疗影像、Re-ID、GAN、NAS、深度估计、自动驾驶、强化学习、车道线检测、模型剪枝&压缩、去噪、去雾、去雨、风格迁移、遥感图像、行为识别、视频理解、图像融合、图像检索、论文投稿&交流、PyTorch和TensorFlow等群。

一定要备注：研究方向+地点+学校/公司+昵称（如论文写作+上海+上交+卡卡），根据格式备注，可更快被通过且邀请进群



▲长按加微信群



▲长按关注CVer公众号



请给CVer一个在看!

[阅读原文](#)