UC Berkeley·CS294-158 | Deep Unsupervised Learning (2020)

CS294 (2020)· 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/BV1uq4y1p7fG

课件 & 代码 · 博客 [扫码或点击链接]
http://blog.showmeai.tech/cs294-158

Berkeley 强化学习 自监督学习 降噪
深度学习 自回归 生成对抗网络 GAN VAE
无监督学习

Awesome AI Courses Notes Cheatsheets 是 **ShowMeAI** 资料库的分支系列，覆盖最具知名度的 **TOP20+** 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！

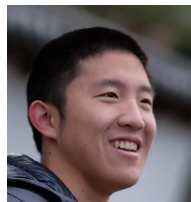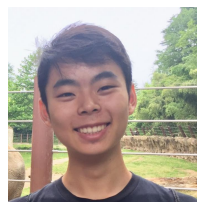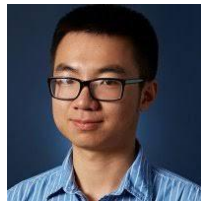| 机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets· 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094

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]

CS294-158 Deep Unsupervised Learning

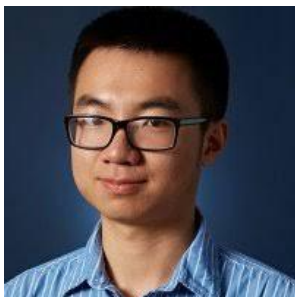## Lecture 1 Intro: Logistics and Motivation

Pieter Abbeel, Xi (Peter) Chen, Jonathan Ho, Aravind Srinivas, Alex Li, Wilson Yan

UC Berkeley

Instructor Team

Pieter Abbeel   Xi (Peter) Chen   Jonathan Ho   Aravind Srinivas   Alex Li   Wilson Yan

Communication

- **Website**: https://sites.google.com/view/berkeley-cs294-158-sp20/home

- **Announcements**
  - Piazza -- sign up today!

- **Questions**
  - Piazza (preferred!)
  - cs294-158-staff@lists.berkeley.edu

- **Office hours**: [all starting *next* week]
  - Pieter: Thu 5-6pm -- 242 Sutardja Dai Hall
  - Alex: Mon 5-6pm, Tue 11-noon -- 326 Soda Hall
  - Wilson: Wed noon-1pm, Fri 2-3pm -- 347 Soda Hall

For homework, TA office hours are the best venue.
For other questions (lecture, final project, research, etc.) any office hours should be great fits

Admission into the Course

- Application: see website

- We'll review applications before the end of the weekend!

Syllabus

Week 1 (1/22) **Intro**

Week 2 (1/29) **Autoregressive Models**

Week 3 (2/5)   **Flow Models**

Week 4 (2/12) **Latent Variable Models**

Week 5 (2/19) **Implicit Models / Generative Adversarial Networks**

Week 6 (2/26) **Implicit Models / Generative Adversarial Networks (ctd) + Final Project Discussion**

Week 7 (3/4)   **Self-Supervised Learning / Non-Generative Representation Learning**

Week 8 (3/11) **Self-Supervised Learning / Non-Generative Representation Learning**

Week 9 (3/18) **Strengths and Weaknesses of Unsupervised Learning Methods**

Spring Break Week (no lecture)

Week 10 (4/1) **Semi-Supervised Learning; Unsupervised Distribution Alignment**

Week 11 (4/8) **Compression**

Week 12 (4/15) **Language Models**

Week 13 (4/22) **Midterm**

Week 14 (4/29) **Representation Learning in Reinforcement Learning**

Week 15 (5/6)   RRR Week (no lecture)

Week 16 (5/13) **Final Project Presentations + Final Project Reports due**

Homework

- HW1: Autoregressive Models (out 1/29, due 2/11)

- HW2: Flow Models (out 2/12, due 2/25)

- HW3: Latent Variable Models (out 2/26, due 3/10)

- HW4: Implicit Models / GANs (out 3/11, due 3/31)

Homework Policy

- **Collaboration:** *Students may discuss assignments. However, each student must code up and write up their solutions independently*.

- **Late assignments:** Recognizing that students may face unusual circumstances and require some flexibility in the course of the semester, each student will have ***a total of 7 free late (calendar) days*** to use as s/he sees fit, but ***no more than 4 late days can be used on any single assignment***. Late days are counted at the granularity of days: e.g., 3 hours late is one late day.

Midterm

■ Date: 4/22 (during lecture slot)

■ Topics: everything covered through (and including) 4/15

■ Format: we will provide a document with questions and answers ahead of time (~20)

■ Rationale: opportunity to force yourself to fully internalize key derivations and concepts

Final Project

- **SCOPE:**
  - Goal: explore and push the boundaries in unsupervised learning.
  - E.g. proposal+evaluation of new algorithms / architectures, investigation of an application of unsupervised learning, benchmarking unsupervised learning, compression, studying synergies between unsupervised learning and other types of learning, etc.
  - Ideally, the project covers interesting new ground and might be the foundation for a future conference paper submission or product.

- **PROJECT TOPICS / STAFF INPUT:**
  - We encourage trying to come up with your own project idea. We are also happy to make suggestions and/or brainstorm ideas together.
  - One of the main reasons we are so excited to teach this class is to see more Deep Unsupervised Learning projects happen. We are very excited to advise on your projects, please don't hesitate to come to office hours to discuss project ideas, project progress, ideas for next steps, etc.

Final Project -- Timeline

- March 2nd Project Proposals Due: 1 page description of project + goals for milestone. -- Submission through google doc shared with instructors, so we can give feedback/suggestions most easily.

- March 9th Approved Project Proposals Due: by this time your proposals should have incorporated instructor feedback, at this stage it should be assured that your proposal is of right fit and scope

- April 13th 3-Page Milestone Due: This is to make sure you are indeed making progress on the project and an opportunity to get feedback on your progress thus far, as well as on any revisions you might want to propose to your project goals. Expectation is that you report on some initial experimental findings (or if you are doing something purely theoretical, some initial progress on that front). -- Submission through google doc shared with instructors, so we can give feedback/suggestions most easily.

- May 13th Project Presentations: 250 SDH, 5-8pm (same as lecture slot)

- May 15th 6-Page Final Project Reports Due

Grading Logistics

- 60% Homework

- 10% Midterm

- 30% Final Project

Do we need to attend class?

- No hard requirement

- BUT: very highly recommended
  - Great opportunity to get to know other students at Berkeley embarking on Deep Unsupervised Learning
  - Pizza!

WARNING

Second offering of this course

There will be some rough edges, please bear with us
+ give feedback!

What is Deep Unsupervised Learning?

- Capturing rich patterns in raw data with deep networks in a **label-free** way
  - Generative Models: recreate raw data distribution
  - Self-supervised Learning: "puzzle" tasks that require semantic understanding
- But why do we care?

Geoffrey Hinton

(in his 2014 AMA on Reddit)

"The brain has about 10^14 synapses and we only live for about 10^9 seconds. So we have a lot more parameters than data. This motivates the idea that we must do a lot of unsupervised learning since the perceptual input (including proprioception) is the only place we can get 10^5 dimensions of constraint per second."

Yann LeCun

Need tremendous amount of information to build machines that have common sense and generalize

[LeCun-20161205-NeurIPS-keynote]

"Pure" Reinforcement Learning (cherry)
- The machine predicts a scalar reward given once in a while.
- **A few bits for some samples**

LeCake

Supervised Learning (icing)
- The machine predicts a category or a few numbers for each input
- Predicting human-supplied data
- **10→10,000 bits per sample**

Unsupervised/Predictive Learning (cake)
- The machine predicts any part of its input for any observed part.
- Predicts future frames in videos
- **Millions of bits per sample**

(Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

"Ideal Intelligence"

"Ideal Intelligence" is all about compression (finding all patterns)

- Finding all patterns = short description of raw data (low Kolmogorov Complexity)

- Shortest code-length = optimal inference (Solomonoff Induction)

- Extensible to optimal action making agents (AIXI)

Aside from theoretical interests

- Deep Unsupervised Learning has many powerful applications

  - Generate novel data

  - Conditional Synthesis Technology (WaveNet, GAN-pix2pix)

  - Compression

  - Improve any downstream task with un(self)supervised pre-training

    - Production level impact: Google Search powered by BERT
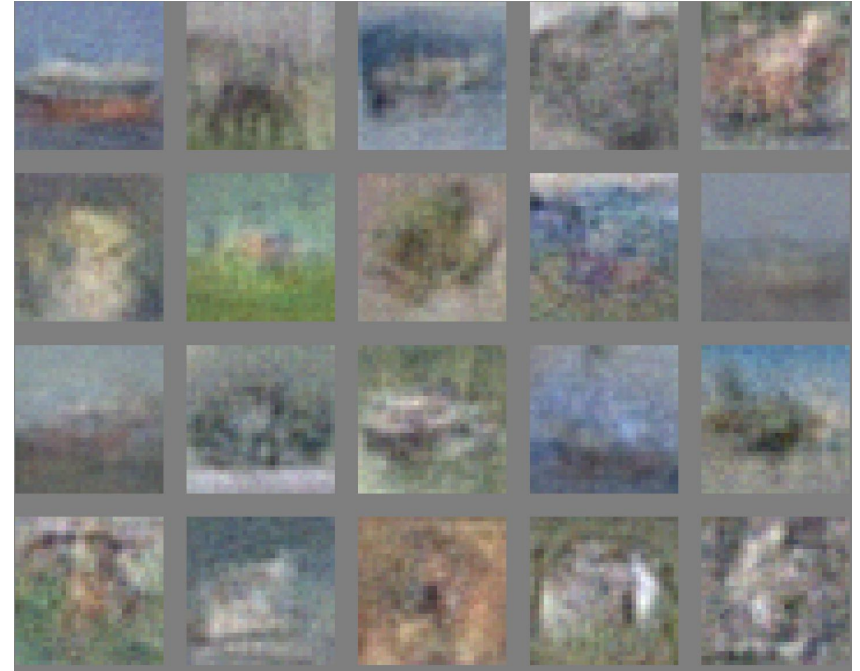
  - Flexible building blocks

Generate Images

[Deep Belief Nets, Hinton, Osindero, Teh, 2006]

Generate Images

[VAE, Kingma and Welling, 2013]

Generate Images

[GAN, Goodfellow et al. 2014]

Generate Images

[DCGAN, Radford, Metz, Chintala 2015]

Generate Images

[DCGAN, Radford, Metz, Chintala 2015]

Generate Images

bicubic
(21.59dB/0.6423)

SRResNet
(23.53dB/0.7832)

SRGAN
(21.15dB/0.6868)

original

[Ledig, Theis, Huszar et al, 2017]

Generate Images

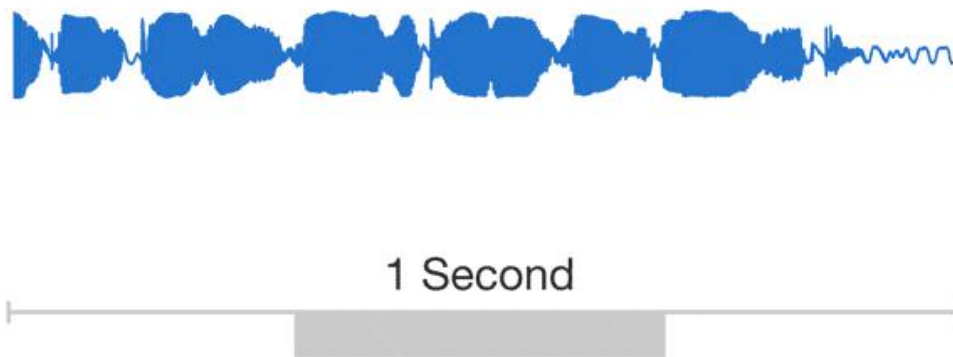[CycleGAN: Zhu, Park, Isola & Efros, 2017]

Generate Images

[BigGAN, Brock, Donahue, Simonyan, 2018]

Generate Images

[StyleGAN, Karras, Laine, Aila, 2018]

Generate Audio

1 Second

Parametric   WaveNet

[WaveNet, Oord et al., 2018]

Generate Video

DVD-GAN: Adversarial Video Generation on Complex Datasets, Clark, Donahue, Simonyan, 2019

Generate Text

PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

[Char-rnn, karpathy, 2015]

Generate Math

\begin{proof}
We may assume that $\mathcal{I}$ is an abelian sheaf on $\mathcal{C}$.
\item Given a morphism $\Delta : \mathcal{F} \to \mathcal{I}$
is an injective and let $\mathfrak q$ be an abelian sheaf on X.
Let $\mathcal{F}$ be a fibered complex. Let $\mathcal{F}$ be a category.
\begin{enumerate}
\item \hyperref[setain-construction-phantom]{Lemma} \label{lemma-characterize-quasi-finite}
Let $\mathcal{F}$ be an abelian quasi-coherent sheaf on $\mathcal{C}$.
Let $\mathcal{F}$ be a coherent $\mathcal{O}_X$-module. Then
$\mathcal{F}$ is an abelian catenary over $\mathcal{C}$.
\item The following are equivalent
\begin{enumerate}
\item $\mathcal{F}$ is an $\mathcal{O}_X$-module.
\end{lemma}

For $\bigoplus_{n=1,\dots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get
$$S = \operatorname{Spec}(R) = U \times_X U \times_X U$$
and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\operatorname{Spec}(R') \to S$ is smooth or an
$$U = \bigcup U_i \times_{S_i} U_i$$
which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\operatorname{GL}_{S'}(x'/S'')$ and we win. $\square$

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that
$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\operatorname{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$
is a unique morphism of algebraic stacks. Note that
$$\text{Arrows} = (Sch/S)_{fppf}^{opp}, (Sch/S)_{fppf}$$
and
$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \operatorname{Spec}(A))$$
is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. $\square$

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,\acute{e}tale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

[Char-rnn, karpathy, 2015]

Generate Text

Prompt [human written]: A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.

GPT-2: *The incident occurred on the downtown train line, which runs from Covington and Ashland stations. In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.*

'The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation,' said Tom Hicks, the U.S. Energy Secretary, in a statement. 'Our top priority is to secure the theft and ensure it doesn't happen again.'

The stolen material was taken from the University of Cincinnati's Research Triangle Park nuclear research site, according to a news release from Department officials.

The Nuclear Regulatory Commission did not immediately release any information. According to the release, the U.S. Department of Energy's Office of Nuclear Material Safety and Security is leading that team's investigation.

'The safety of people, the environment and the nation's nuclear stockpile is our highest priority,' Hicks said. 'We will get to the bottom of this and make no excuses.'

[OpenAI's GPT2 -- Radford, Wu, Child, Luan, Amodei, Sutskever, 2019]

Generate Text

Prompt (human written): Recycling is good for the world, no, you could not be more wrong.

GPT-2: Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world's most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources.

[OpenAI's GPT2 -- Radford, Wu, Child, Luan, Amodei, Sutskever, 2019]

Try it yourself

https://talktotransformer.com/

Compression - Lossless

Model	Bits per byte
CIFAR-10	
PixelCNN (Oord et al., 2016)	3.03
PixelCNN++ (Salimans et al., 2017)	2.92
Image Transformer (Parmar et al., 2018)	2.90
PixelSNAIL (Chen et al., 2017)	2.85
Sparse Transformer 59M (strided)	**2.80**
Enwik8	
Deeper Self-Attention (Al-Rfou et al., 2018)	1.06
Transformer-XL 88M (Dai et al., 2018)	1.03
Transformer-XL 277M (Dai et al., 2018)	**0.99**
Sparse Transformer 95M (fixed)	**0.99**
ImageNet 64x64	
PixelCNN (Oord et al., 2016)	3.57
Parallel Multiscale (Reed et al., 2017)	3.7
Glow (Kingma & Dhariwal, 2018)	3.81
SPN 150M (Menick & Kalchbrenner, 2018)	3.52
Sparse Transformer 152M (strided)	**3.44**
Classical music, 5 seconds at 12 kHz	
Sparse Transformer 152M (strided)	**1.97**

Generative models provide better bit-rates than distribution-unaware compression methods like JPEG, etc.

Compression - Lossy

JPEG   JPEG2000   WaveOne

[Rippel & Bourdev, 2017]

Downstream Task - Sentiment Detection

This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

[Radford et al., 2017]

Downstream Tasks - NLP (BERT Revolution)

Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI
1	T5 Team - Google	T5	⬀	90.3	71.6	97.5	92.8/90.4	93.1/92.8	75.1/90.6	92.2	91.9	96.9	92.8	94.5
2	ERNIE Team - Baidu	ERNIE	⬀	90.0	72.2	97.5	93.0/90.7	92.9/92.5	75.2/90.8	91.2	90.8	96.0	90.9	94.5
3	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART	⬀	89.9	69.5	97.5	93.7/91.6	92.9/92.5	73.9/90.2	91.0	90.8	99.2	89.7	94.5
4	王玮	ALICE v2 large ensemble (Alibaba DAMO NLP)	⬀	89.7	73.2	97.1	93.9/91.9	93.0/92.5	74.8/91.0	90.8	90.6	95.9	87.4	94.5
5	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)	⬀	88.4	68.0	96.8	93.1/90.8	92.3/92.1	74.8/90.3	91.1	90.7	95.6	88.7	89.0
6	Junjie Yang	HIRE-RoBERTa	⬀	88.3	68.6	97.1	93.0/90.7	92.4/92.0	74.3/90.2	90.7	90.4	95.5	87.9	89.0
7	Facebook AI	RoBERTa	⬀	88.1	67.8	96.7	92.3/89.8	92.2/91.9	74.3/90.2	90.8	90.2	95.4	88.2	89.0
8	Microsoft D365 AI & MSR AI	MT-DNN-ensemble	⬀	87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9	87.9	87.4	96.0	86.3	89.0
9	GLUE Human Baselines	GLUE Human Baselines	⬀	87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9
10	Stanford Hazy Research	Snorkel MeTaL	⬀	83.2	63.8	96.2	91.5/88.5	90.1/89.7	73.1/89.9	87.6	87.2	93.9	80.9	65.1

[https://gluebenchmark.com/leaderboard]

Downstream Tasks - Vision (Contrastive)

Method	Architecture	mAP
Transfer from labeled data:		
Supervised baseline	ResNet-152	74.7
Transfer from unlabeled data:		
Exemplar [17] by [13]	ResNet-101	60.9
Motion Segmentation [47] by [13]	ResNet-101	61.1
Colorization [64] by [13]	ResNet-101	65.5
Relative Position [14] by [13]	ResNet-101	66.8
Multi-task [13]	ResNet-101	70.5
Instance Discrimination [60]	ResNet-50	65.4
Deep Cluster [7]	VGG-16	65.9
Deeper Cluster [8]	VGG-16	67.8
Local Aggregation [66]	ResNet-50	69.1
Momentum Contrast [25]	ResNet-50	74.9
Faster-RCNN trained on CPC v2	ResNet-161	**76.6**

"If, by the first day of autumn (Sept 23) of 2015, a method will exist that can match or beat the performance of R-CNN on Pascal VOC detection, without the use of any extra, human annotations (e.g. ImageNet) as pre-training, Mr. Malik promises to buy Mr. Efros one (1) gelato (2 scoops: one chocolate, one vanilla)."

Table: Data-Efficient Image Recognition using CPC (Henaff, Srinivas, et al)

Summary

- **Unsupervised Learning**: Rapidly advancing field thanks to compute; deep learning engineering practices; datasets; lot of people working on it.
- **Not just an academic interest topic.** Production level impact [example: BERT is in use for Google Search and Assistant].
- **What is true now may not be true even a year from now** [example: self-supervised pre-training was way worse than supervised in computer vision tasks like detection/segmentation last year. Now it is better].
- **Language Modeling** (GPT), **Image Generation** (conditional GANs), **Language pre-training** (BERT), **vision pre-training** (CPC / MoCo) starting to work really well. Good time to learn these well and make very impactful contributions.
- **Autoregressive Density Modeling, Flows, VAEs, UL for RL**, etc have *huge room for improvement*. Great time to work on them.

UC Berkeley·CS294-158 | Deep Unsupervised Learning (2020)

CS294 (2020)· 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/BV1uq4y1p7fG

课件 & 代码 · 博客 [扫码或点击链接]
http://blog.showmeai.tech/cs294-158

Berkeley 强化学习 自监督学习 降噪
深度学习 自回归 生成对抗网络
无监督学习 GAN VAE

Awesome AI Courses Notes Cheatsheets 是 **ShowMeAI** 资料库的分支系列，覆盖最具知名度的 **TOP20+** 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]