Recitation

K-means Crash Course & Large Scale Recognition







Today's agenda

- K-means clustering for A3
- Large Scale Recognition





So far: Segmentation and clustering

• Goal: identify groups of pixels that go together









- These pixel values show that there are three things in the image.
- We could label every pixel in the image according to which of these primary intensities it is.
 - \circ i.e., segment the image based on the intensity feature.
- What if the image isn't quite so simple?

Recitation 7-4





May 16, 2025

Joshua Jung



- How do we determine the three main intensities that define our groups?
- Assumption: each cluster has a cluster center
 - A mean cluster value.





- Goal: choose three "centers" as the representative intensities and label every pixel according to which of these centers it is nearest to.
- Best cluster centers are those that minimize Sum of Square Distance (SSD) between all points and their nearest cluster center c_i:

$$SSD = \sum_{C} \sum_{v \in C} (v - c_i)^2$$

Recitation 7-7

Clustering

• With this difficult objective,

 If we knew the *cluster centers*, we could allocate points to groups by assigning each to its closest center.



 If we knew the group memberships, we could get the centers by computing the mean per group.



Recitation 7-8



Given, a set of points, randomly select k=3 of them to be the cluster centers



Joshua Jung

Recitation 7-9

Categorize each point into a cluster defined by its closest center.

Next, move the cluster centers to the mean location amonast its cluster



Recitation 7-10

Joshua Jung

Repeat with new cluster center locations



Joshua Jung

Recitation 7-11

Categorize into new clusters. Move center to the mean



Joshua Jung

Recitation 7-12

Repeat with new cluster centers



Joshua Jung

Recitation 7-13

Large Scale Recognition







California Roll

Ingredients: Rice, Seaweed, Crab, Cucumber, Avocado Calories: 40 Fat: 7g Carb: 40g Protein: 5g Gluten Free

Joshua Jung





Recitation 7 16



Mountain Lion DO NOT RUN Raise arms to appear larger.

Show your teeth







Recitation 7 18



Recitation 7 19

I wish my computer could recognize <u>EVERYTHING</u>



May 16, 2025

Joshua Jung



Surveillance



Robotics



Wearable devices



Smart photo album



Image search



Assistive tools



Driverless cars



Mining social media

Joshua Jung

Recitation 7 21

What can computers already recognize?



















Recitation 7-24







What's the next to work on? Coffee Mugs!











No close image matches found

Avoid glare from the flash.
Zoom in as much as possible by placing your device close to whatever you want to photograph.

-

Joshua Jung

Recitation 7 28

PASCAL VOC [Everingham et al. 2006-2012]



Airplane	Dining table
Bird	Dog
Boat	Horse
Bike	Motorbike
Bottle	Person
Bus	Potted plant
Car	Sheep
Cat	Sofa
Chair	Train
Cow	TV monitor

No Coffee Mugs!

Joshua Jung

Recitation 7 29

The rest of the talk will be about **Coffee Mugs**!







The rest of the talk will be about Coffee Mugs



Joshua Jung

Recitation 7 31

Let's work on recognizing **EVERYTHING**







How many things are there?



Joshua Jung

Recitation 7 33

From 20 classes to Millions?

Joshua Ju





May 16, 2025

Joshua Jung

Big Data from the Internet





Global Consumer Internet Traffic Per Month









Big Data from the Internet

□ The Internet can teach EVERYTHING









Evolution Gone Wild Future plants and animals

http://www.worth1000.com/contests/12705/contest

Joshua Jung





What kind of credit card is President Obama using in this video of him donating to his campaign?

Joshua Jung

Recitation 7 41



Teach machines to recognize **EVERYTHING**

Joshua Ju	ung
-----------	-----

Recitation 7 42



Goal: Build a recognition engine on EVERYTHING 10K classes





IMAGENET [Deng et al. 2009]

22K categories and **14M** images

- Animals
 - Bird
 - Fish
 - Mammal
 - Invertebrate

- Plants
 Structures
 - Tree
 Artifact
- Flower
 Tools
 - Food
 - Appliances
 - Materials
 Structures

- Person
- Scenes
 - Indoor
 - Geological Formations

www.image-net.org

Sport Activities

Joshua Ju





Recitation 7

May 10, 2024

Learn to Classify 10K Classes

Model	Time	Accuracy	Num. of Parameters	Num. of FLOPs	Num. of Layers
AlexNet [21]	2012	57.2%	60 M	720 M	8
VGGNet [22]	2014	71.5%	138 M	15,300 M	16
GoogleNet [23]	2014	69.8%	6.8 M	1,500 M	22
ResNet [24]	2015	78.6%	55 M	2,300 M	152
DenseNet [25]	2017	79.2%	25.6 M	1,150 M	190
SENet [26]	2017	82.7%	145.8 M	42,300 M	-
NASNet [27]	2018	82.7%	88.9 M	23,800 M	-
SqueezeNet [29]	2016	57.5%	1.2 M	833 M	-
MobileNet [30]	2017	70.6%	4.2 M	569 M	28
ShuffleNet [31]	2018	73.7%	4.7 M	524 M	-
ShiftNet-A [32]	2018	70.1%	4.1 M	1,400 M	<u></u>
FE-Net [33]	2019	75.0%	5.9 M	563 M	-



Joshua Jung

Recitation 7 46

Still not everything!

- ImageNet still doesn't contain everything!
 - \circ Designed in 2009
 - Wouldn't be able to identify modern concepts!
 - Airpod
 - OpenAl
 - Lacks flexibility real-word categories are constantly evolving, the dataset doesn't
 - costly to classify more images
 - \circ Biases
 - Data comes mainly from Western-centric sources
 - Images are high-quality images and doesn't reflect real-world conditions (weather, certain contexts, etc).

Joshua Jung

Recitation 7-47

Transfer Learning

- What is Transfer learning?
 - Leverage a model pre-trained on a larger dataset (ImageNet) to solve a different, but related task
 - Utilizes learned features (e.g. edges, textures) that are generic across vision tasks
- Why use ImageNet pre-trained models?
 - Efficient data usage
 - Robust feature extraction
 - Reduced training time
 - Improved performance



Joshua Jung

Recitation 7-48

Transfer Learning

- Approaches to transfer learning
 - Feature Extraction
 - Freeze the early layers (generic features) to use as the backbone for the new model
 - Train a classifier head to use the backbone in the specific task
 - \circ Fine Tuning
 - Start with pretrained weights
 - Freeze the weights of the big model and just train a few layers to "fine tune" to the new dataset

Joshua Jung

Recitation 7-49

Let's test it out!

We now have big VLM's like ChatGPT!





Moringa (Japanese flying squirrel)



What do you see in this picture

The image shows a group of small, fluffy animals perched on a tree branch. They appear to be **Japanese dwarf flying squirrels**—recognizable by their large, dark eyes, soft gray-brown fur, and round faces. These squirrels are known for their gliding ability, enabled by a membrane that stretches between their limbs. They are typically found in Japan and are known for their incredibly cute, wide-eyed appearance.

Joshua Jung

ChatGPT 40 ~

Recitation 7-51

Amanita phalloides



Joshua Jung

Recitation 7-52





Recitation 7-53

Geoguessr



Where is this located

Thought for 27 seconds >

It looks like a quiet two-lane road winding through dense conifer forest—spruce and fir on both sides—with a neatly mown grassy verge and a solid double-white centerline. Those pavement markings, together with the mix of tall pines, low birches, and an orchard-style cluster of smaller trees on the right, are much more typical of Northern European rural highways than of North American ones (in the U S. you'd usually see a yellow centerline on a two-way road).

So while I can't pin down an exact spot from the photo alone, the scene matches what you'd see along secondary roads in southern Finland, Sweden, or the Baltic states—flat terrain, boreal forest, and that distinct Nordic road striping. If you took the picture yourself, any nearby signage, kilometer posts, or the language on warning signs would



Joshua Jung

Recitation 7-54





Recitation 7-55

Geoguessr



Where is this located?

Thought for 2m 48s >

The terrain, road furniture, and small details (like the triangular-over-square roadside sign and the scattered wind turbines on an otherwise perfectly flat plain) match what you see all over the lower Po Valley in northern Italy. In other words, this looks very much like the open farmland south-east of Ferrara/Bologna in Emilia-Romagna.

ወ ይ ወ ወ ወ ሪ~

Joshua Jung

Analyzed image



I see what might be a white statue of Mary, which is typical of Italy. The autostrada guardrails
and triangular signs also point toward Italy. However, the triangular sign might be hard to see,
possibly appearing as a silhouette in black. It could even be a European hazard sign. In the
horizon, there's a building resembling a bell tower or a silo. I'll zoom a bit to the left of center
to get a better look at that structure.



Recitation 7-56





Recitation 7-57

Geoguessr



Where is this picture from?

Thought for 3m 32s >

Looks like Seattle's South Lake Union tech district—specifically the intersection of **Westlake Avenue N and Mercer Street**, right by the Allen Institute building (with the green-and-orange frames) and a Starbucks Reserve at street-level in one of Amazon's campus towers.

linked to the Starbucks Reserve line. Let me confirm and adjust further.

 It looks like the building might be an Amazon structure with a Starbucks Reserve storefront. The building's bright green accent on the right side could be typical of Amazon's Doppler building or Seattle's South Lake Union area near Westlake Avenue. But I'm wondering if the building with the green accent could be the Allen Institute for Brain Science. The backdrop features what seems like a cylindrical high-rise, possibly the 400 Fairview or Washington State Research building. Let me inspect the larger view to confirm.



Joshua Jung

Recitation 7-58