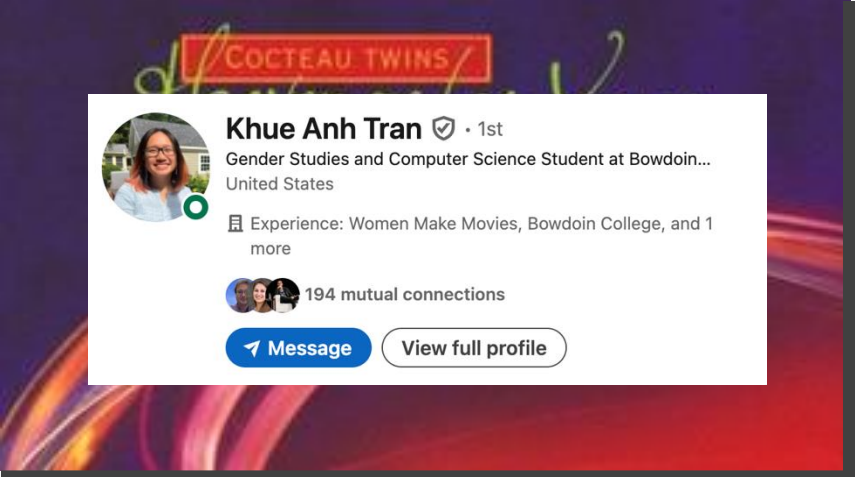
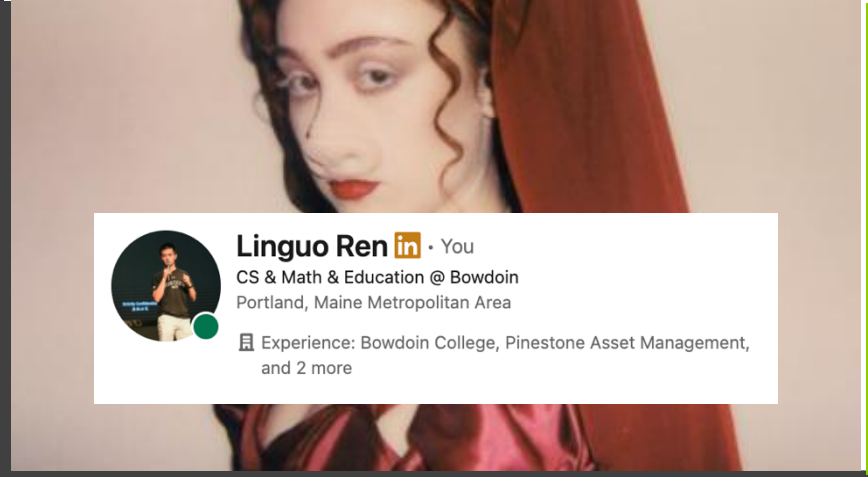
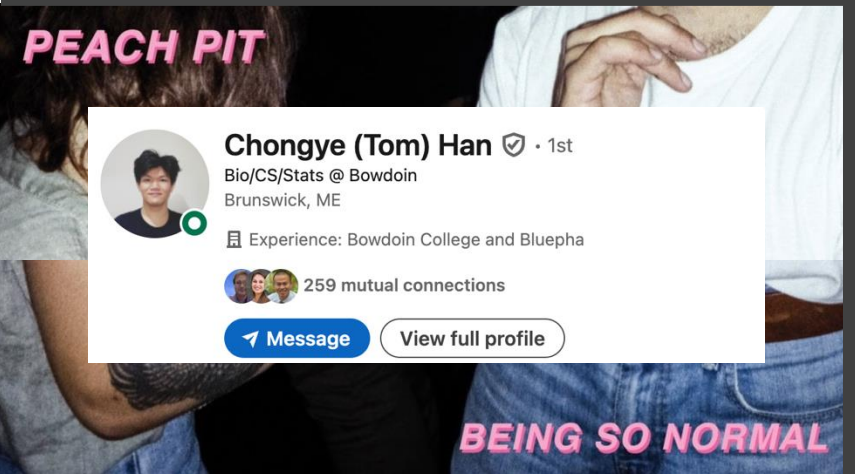






album-wiz
Vinyl Companion

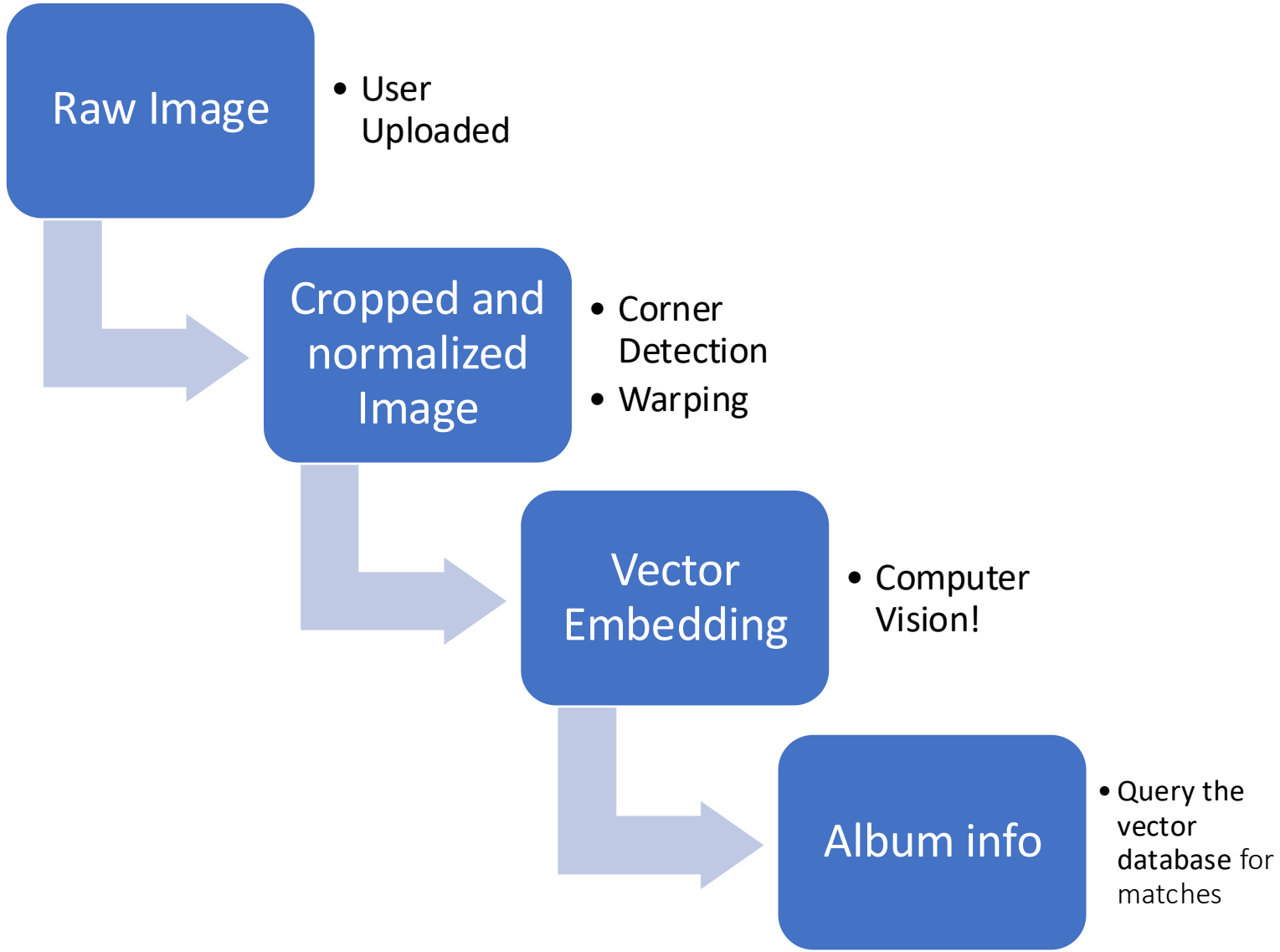
Khue Anh Tran, Mason Daugherty, Tom Han, Linguo Ren

CSCI 3485 *Deep Learning for Computer Vision*

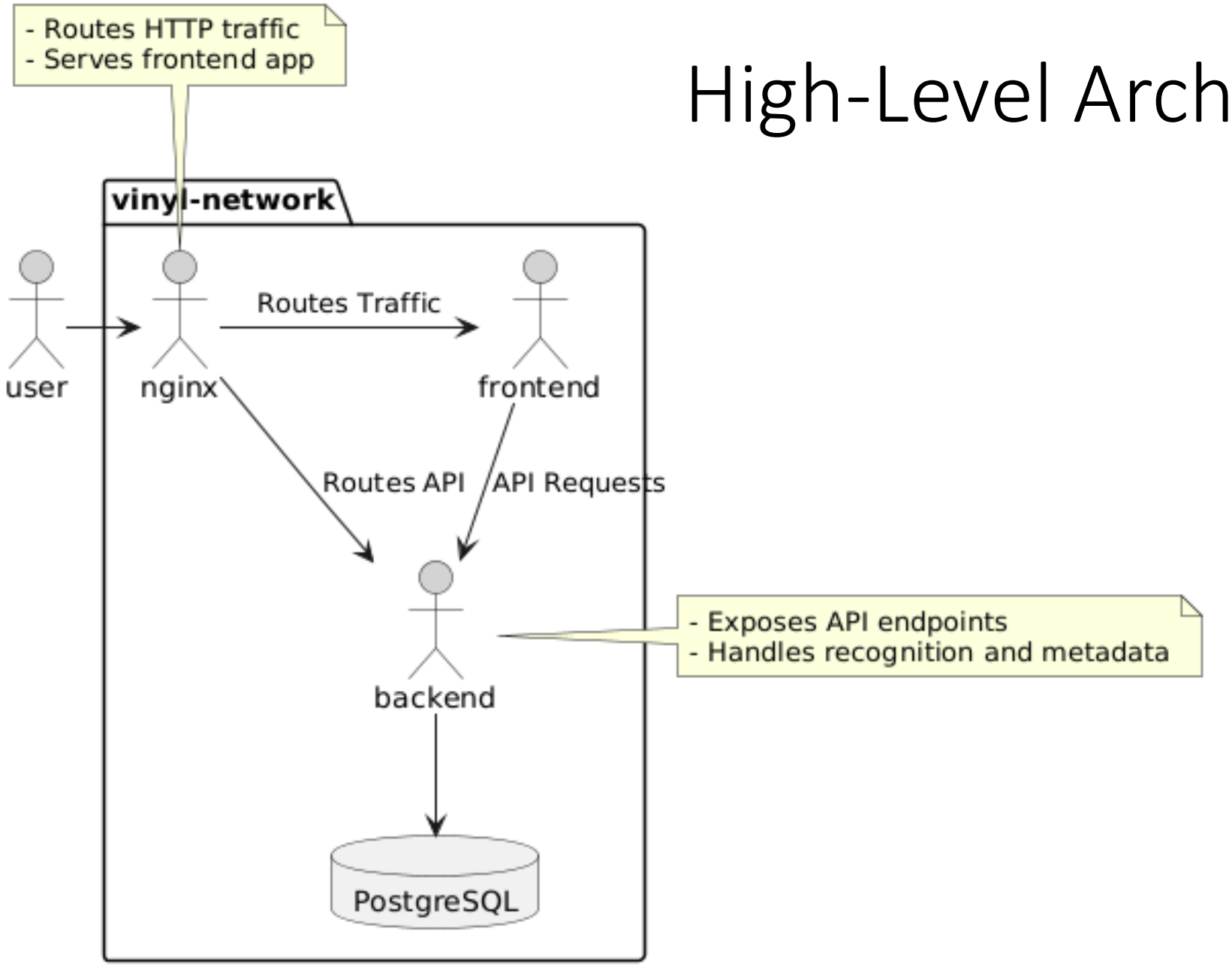
Professor Farias

Dec 17th, 2024


BOWDOIN
'63-'64
GLEE CLUB
tone dogs
THE SMITHS
Bartók, Busoni, Messiaen, Stravinsky
Paul Jacobs, piano
Betty Dylan


Raw Image
• User Uploaded
Cropped and normalized Image
• Corner Detection
• Warping
Vector Embedding
• Computer Vision!
Album info
• Query the vector database for matches

High-Level Architecture

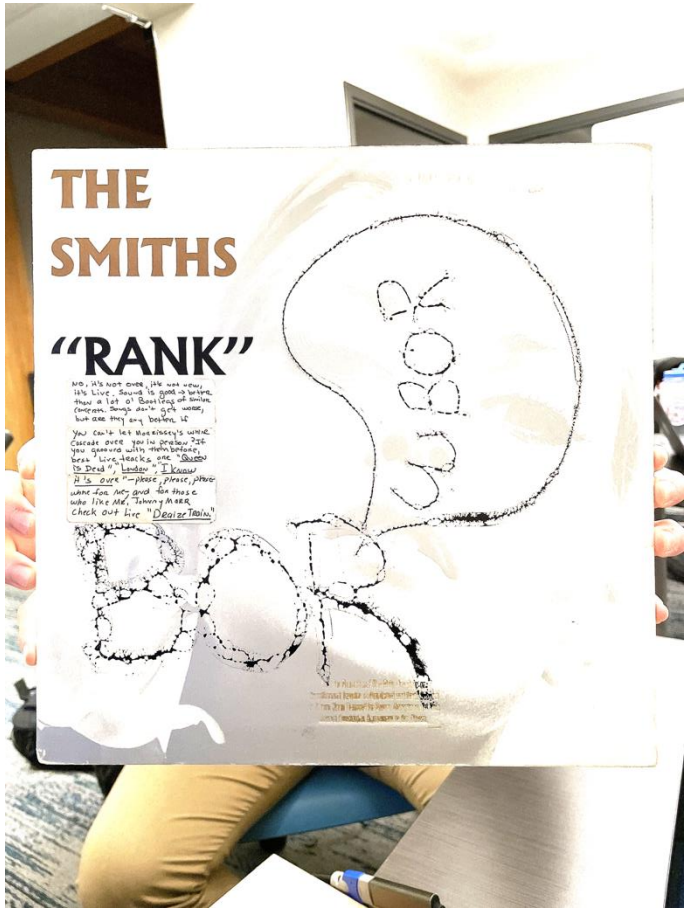


Square (Album) Detection

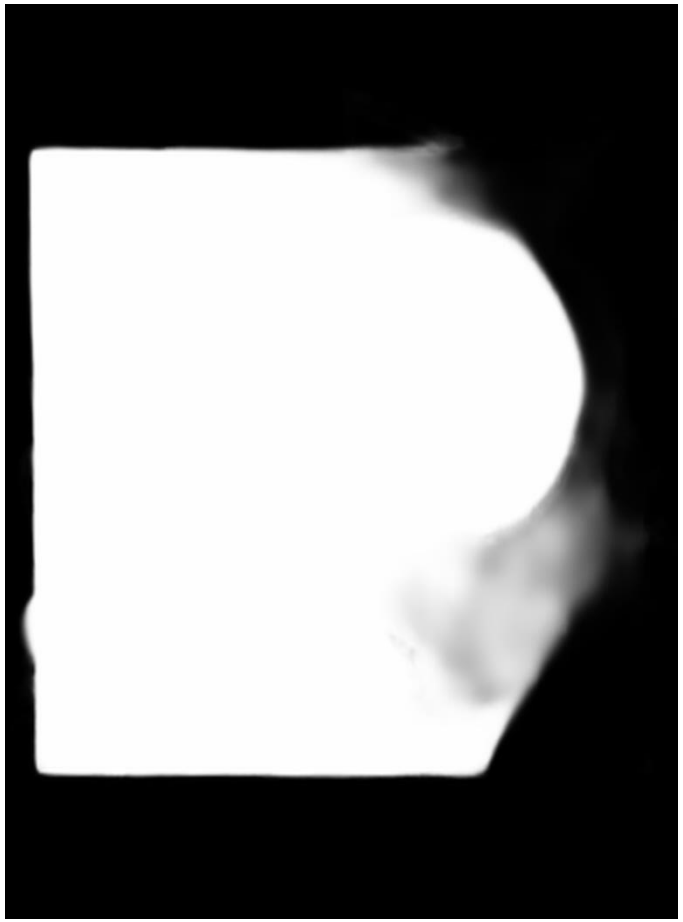
- Image Sharpening
- Background Removal
- Line Detection
- Corner Detection
- Perspective Warping

Original Images

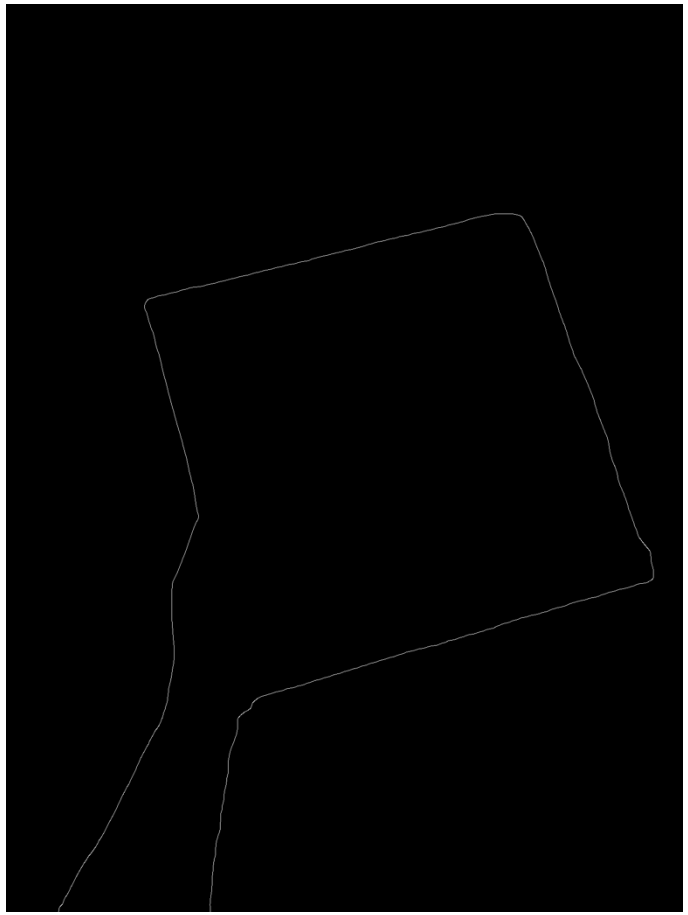
Image Sharpening (help detect edges)

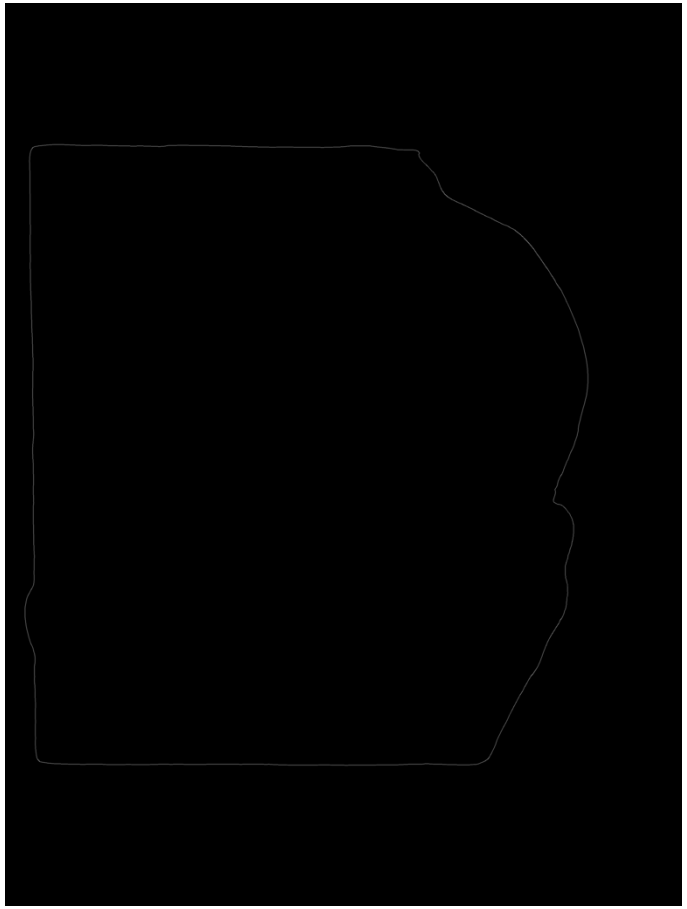


Background Removal (with dilation)

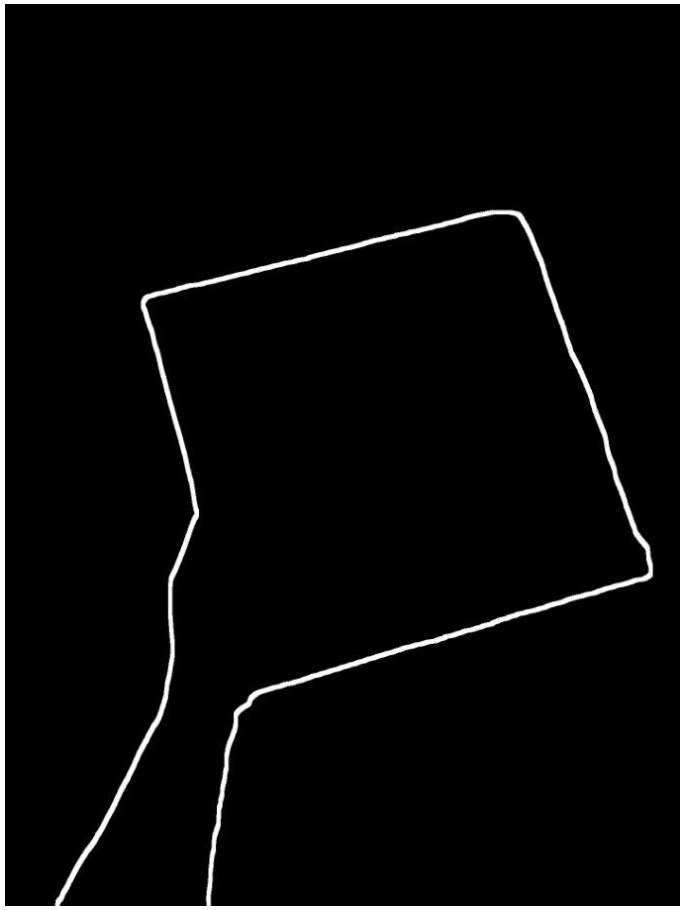


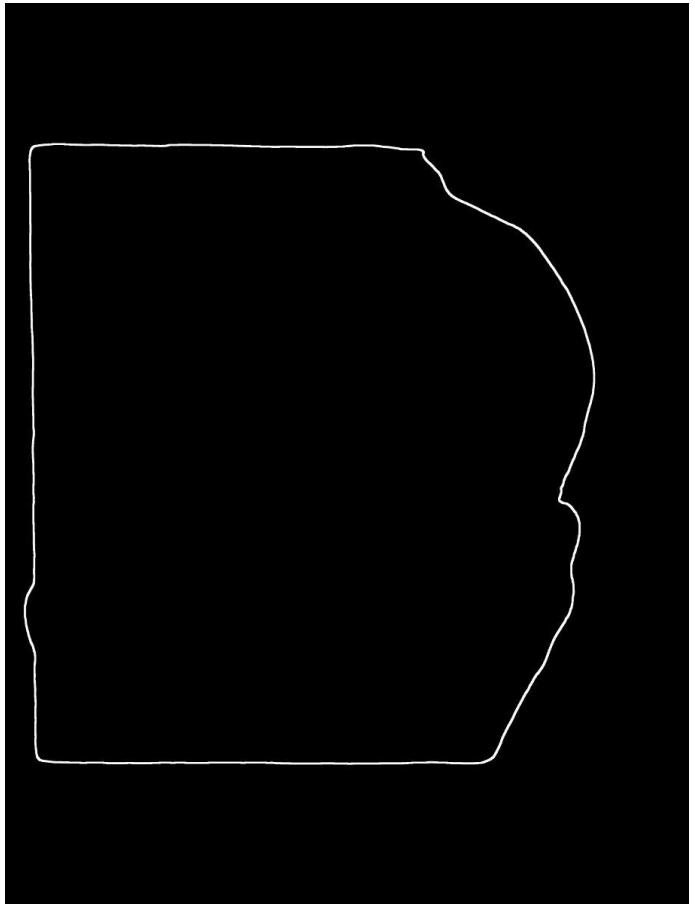
Line Detection: Canny Edges



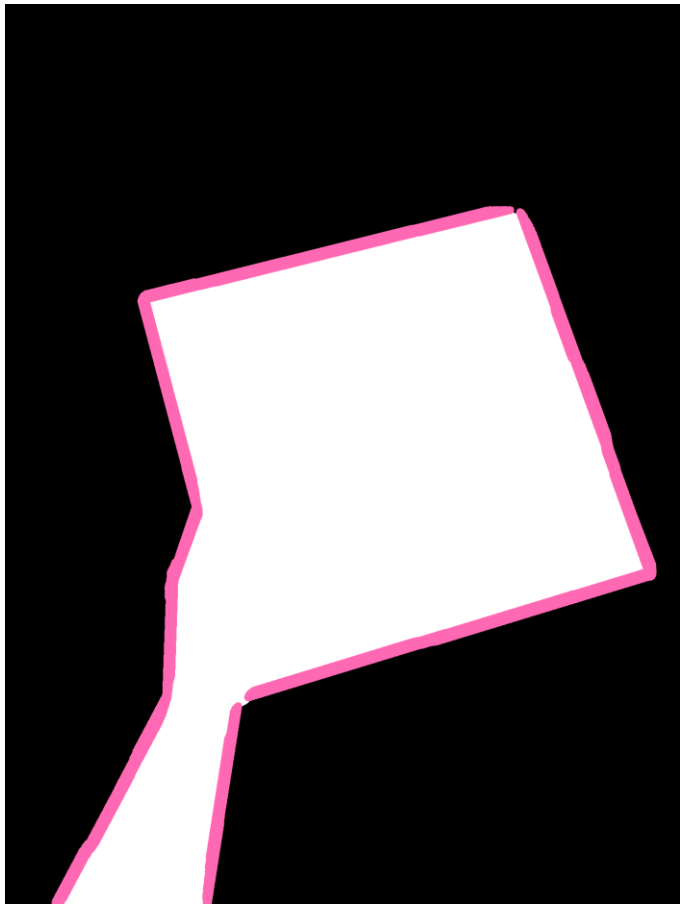


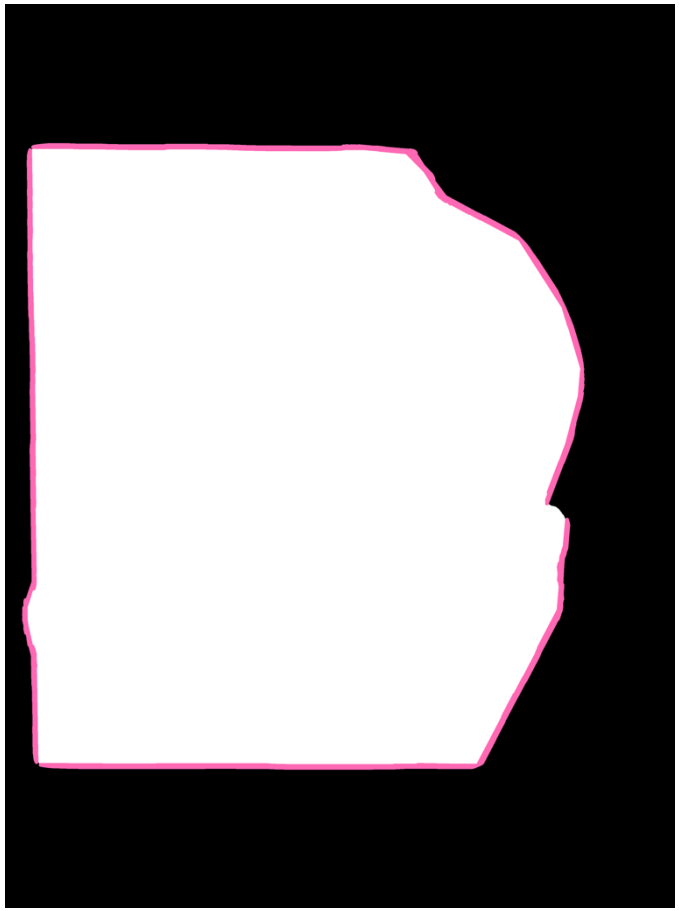
Line Detection: Dilated Edges



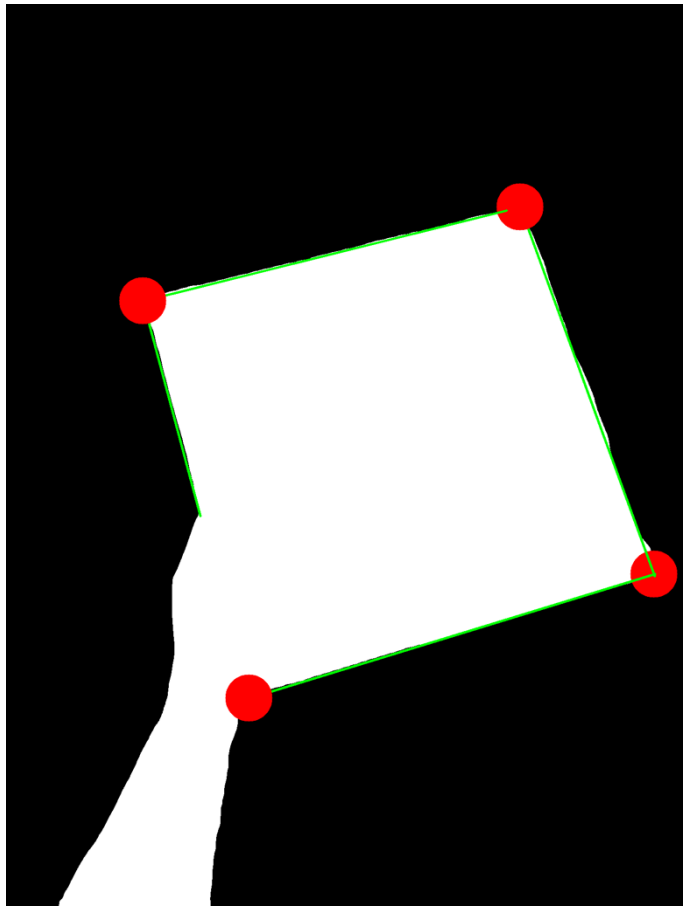


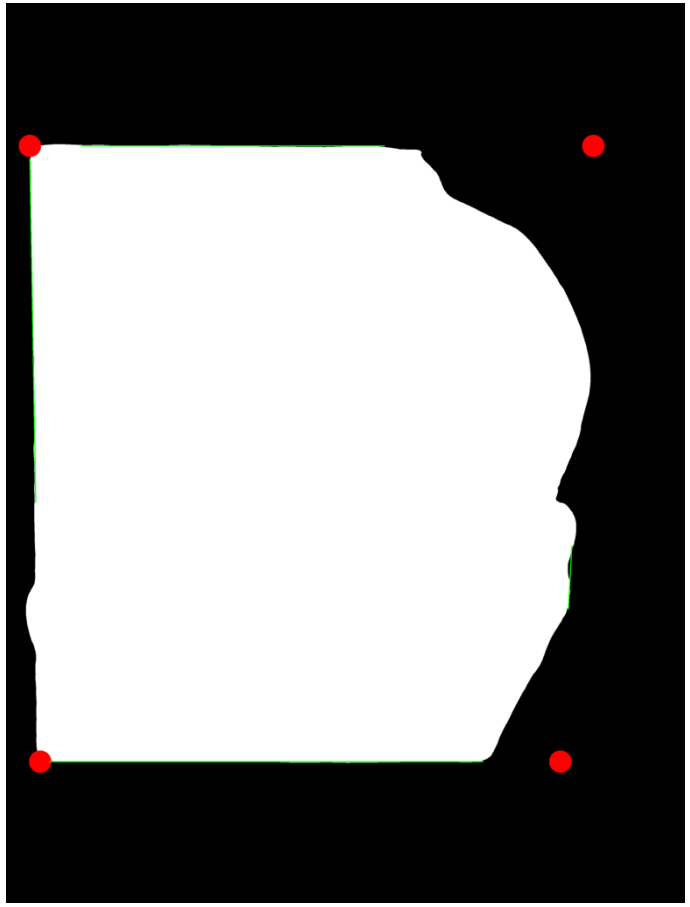
Line Detection: Unique Lines



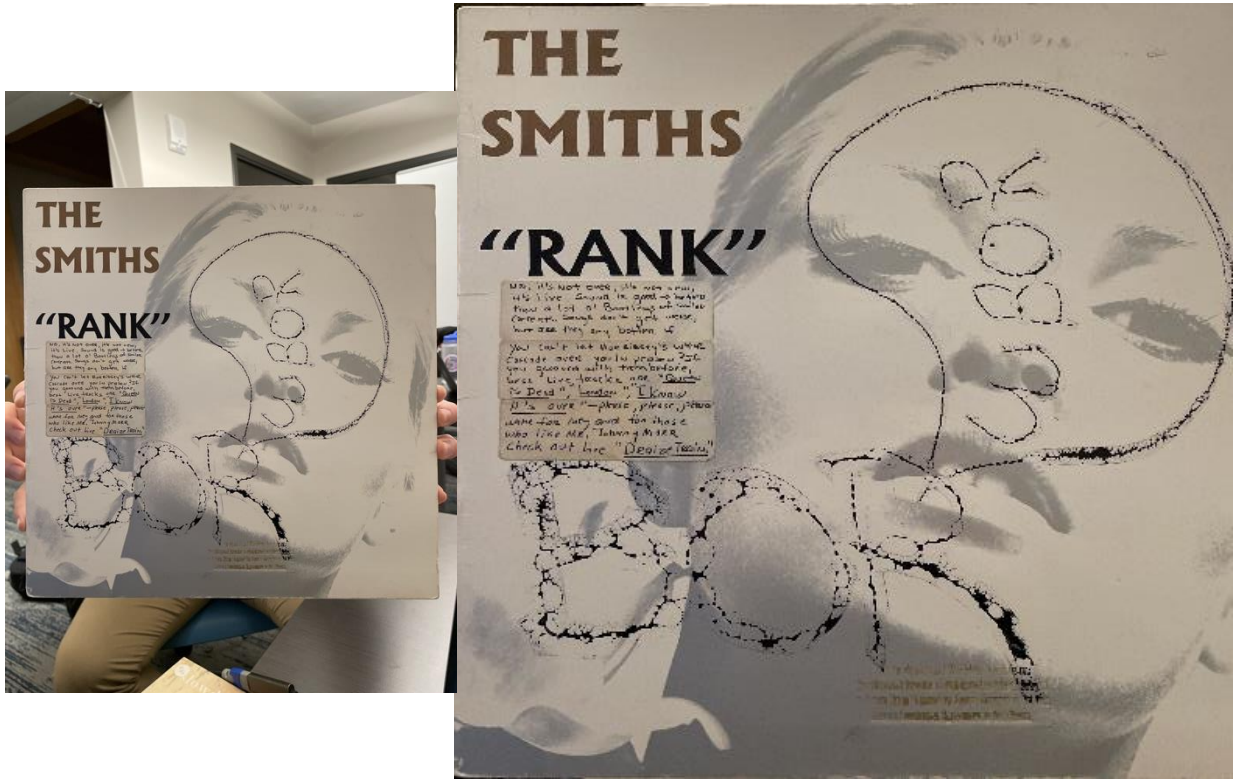


Corner Detection



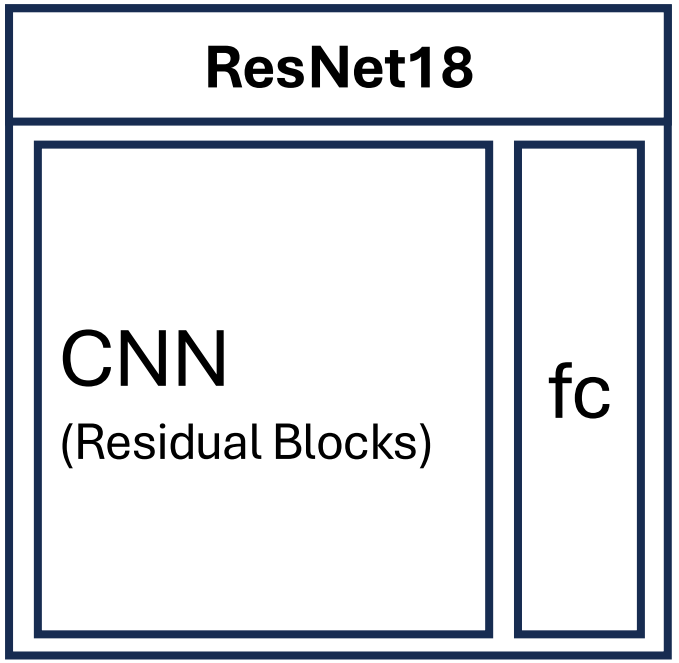


Perspective Warping



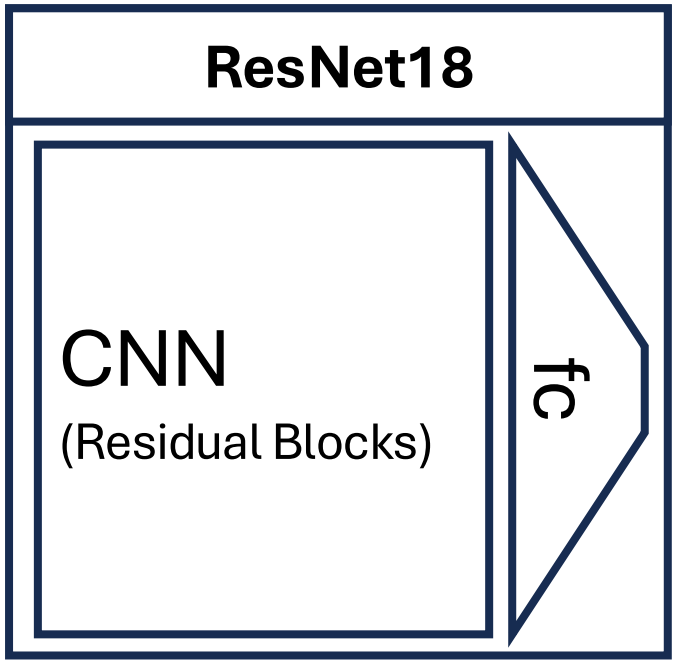
Model Training

- Faster (compared to ResNet51)
- Original output of ResNet18: 1000
  - Too long, not efficient for vector search
  - Not power of 2

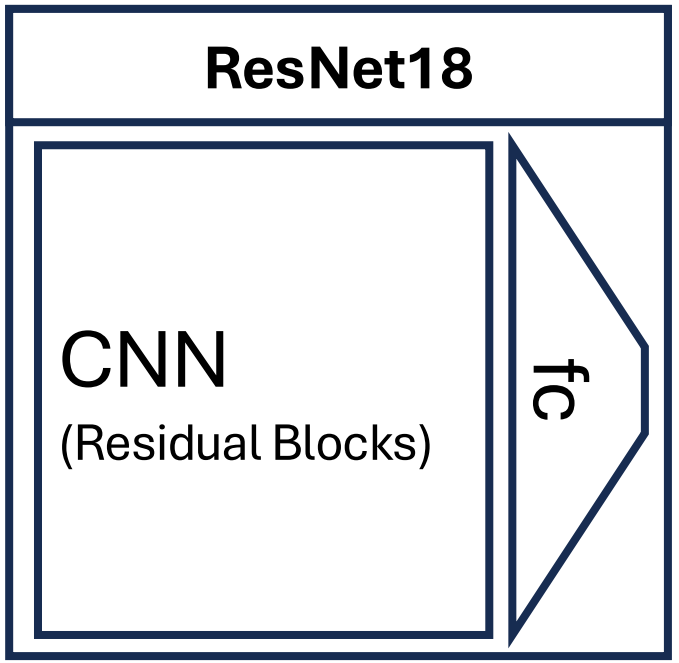


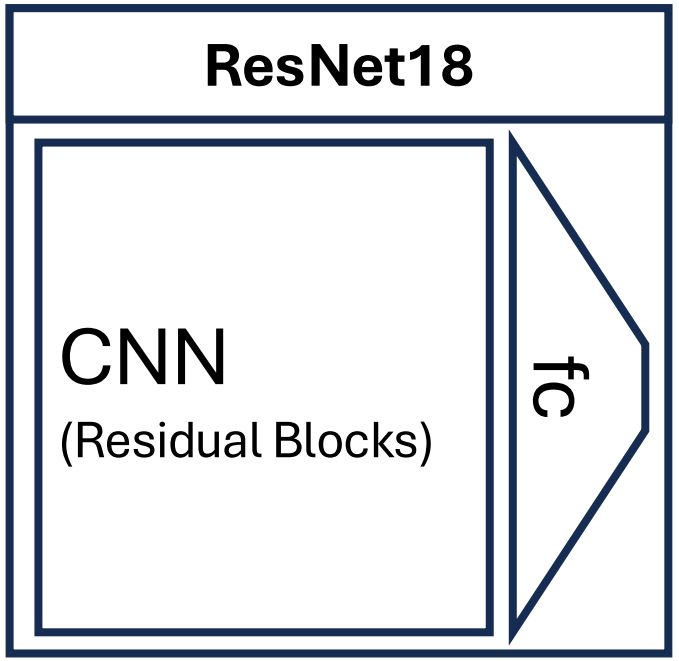
Model Training

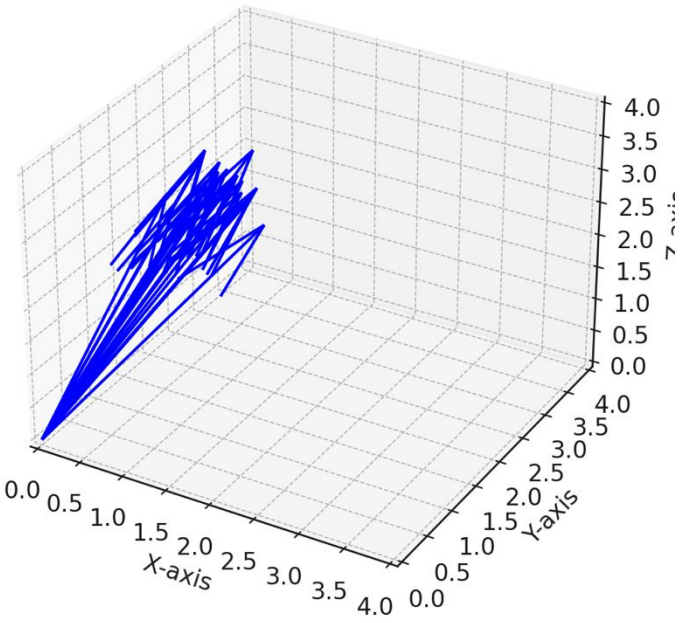
- Original output of ResNet18: 1000
  - Too long, not efficient for vector search
  - Not power of 2
- New length: 256
  - Short
  - Is power of 2
  - BONUS: Faster to train!
  - SKINNY

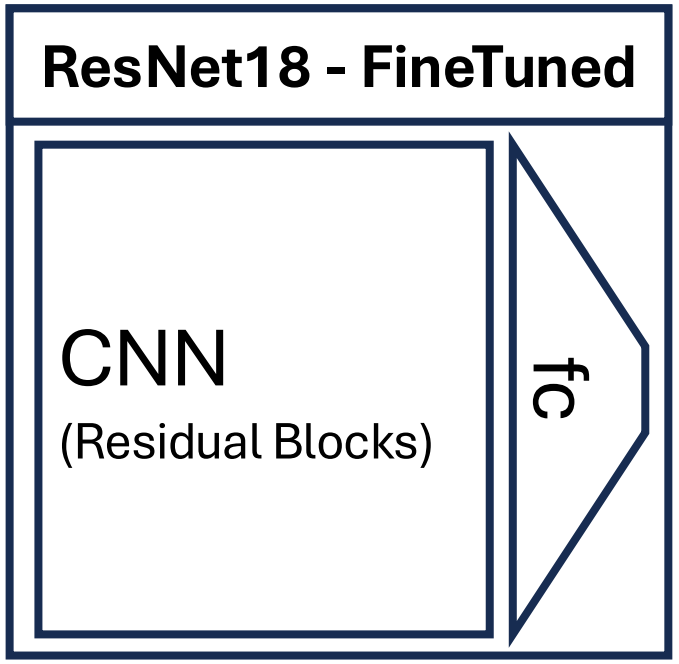


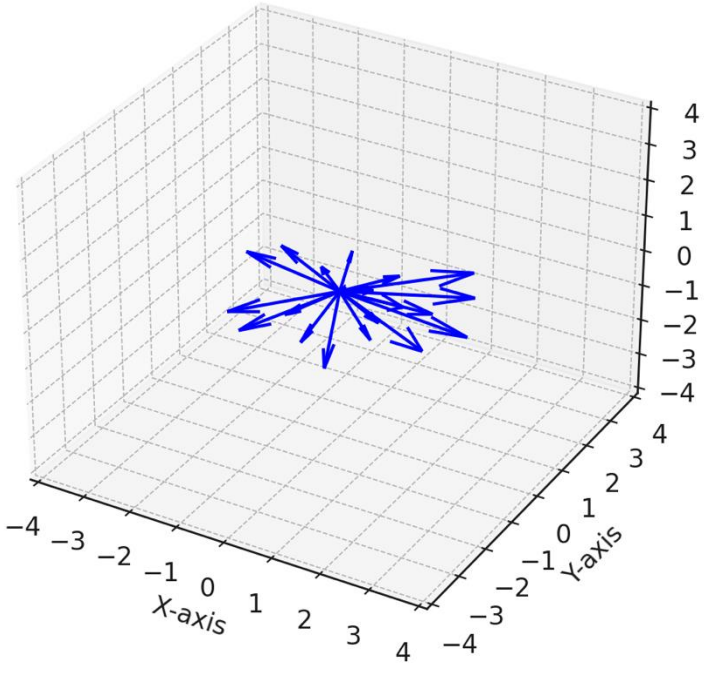

GALAXIE 500
LANA DEL REY
BORN TO DIE
The Paradise Edition
PEACH PIT
brat
BEING SO NORMAL

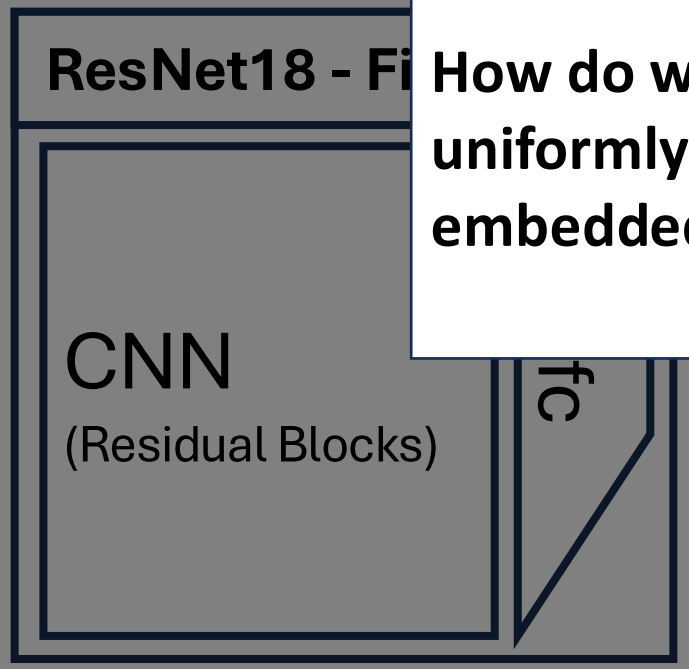

ResNet18
CNN
(Residual Blocks)
fc


ResNet18
CNN
(Residual Blocks)
fc

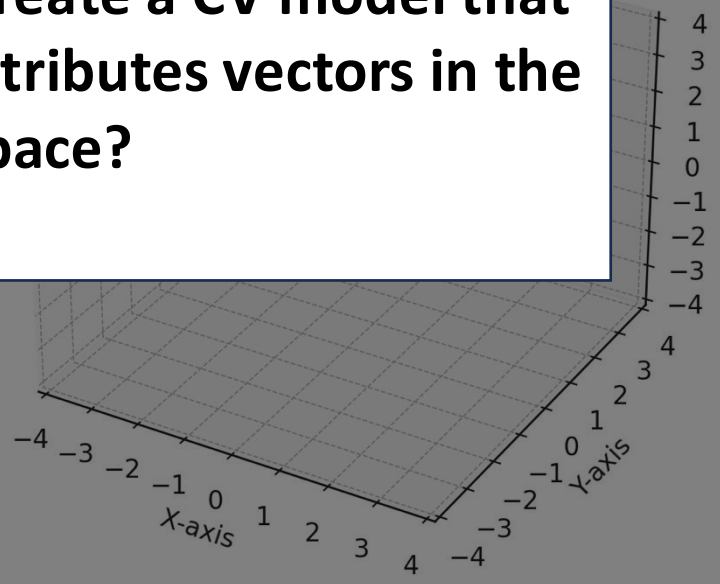

X-axis
Y-axis
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

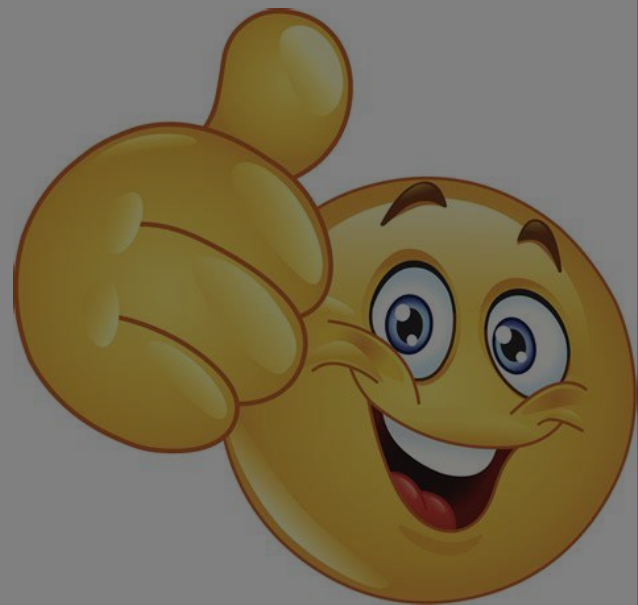

ResNet18 - FineTuned
CNN
(Residual Blocks)
fc


X-axis
Y-axis



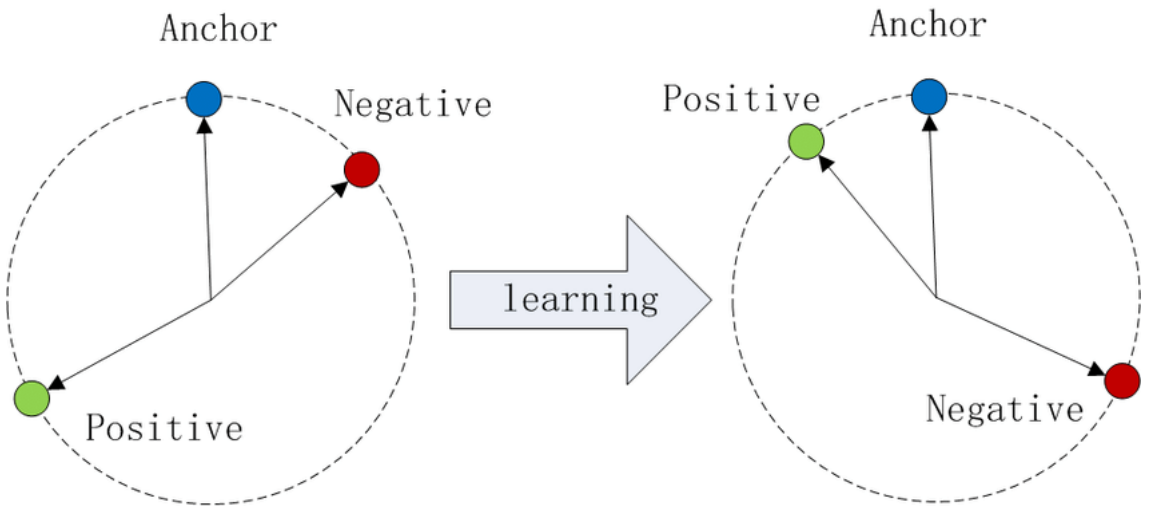
How do we create a CV model that uniformly distributes vectors in the embedded space?



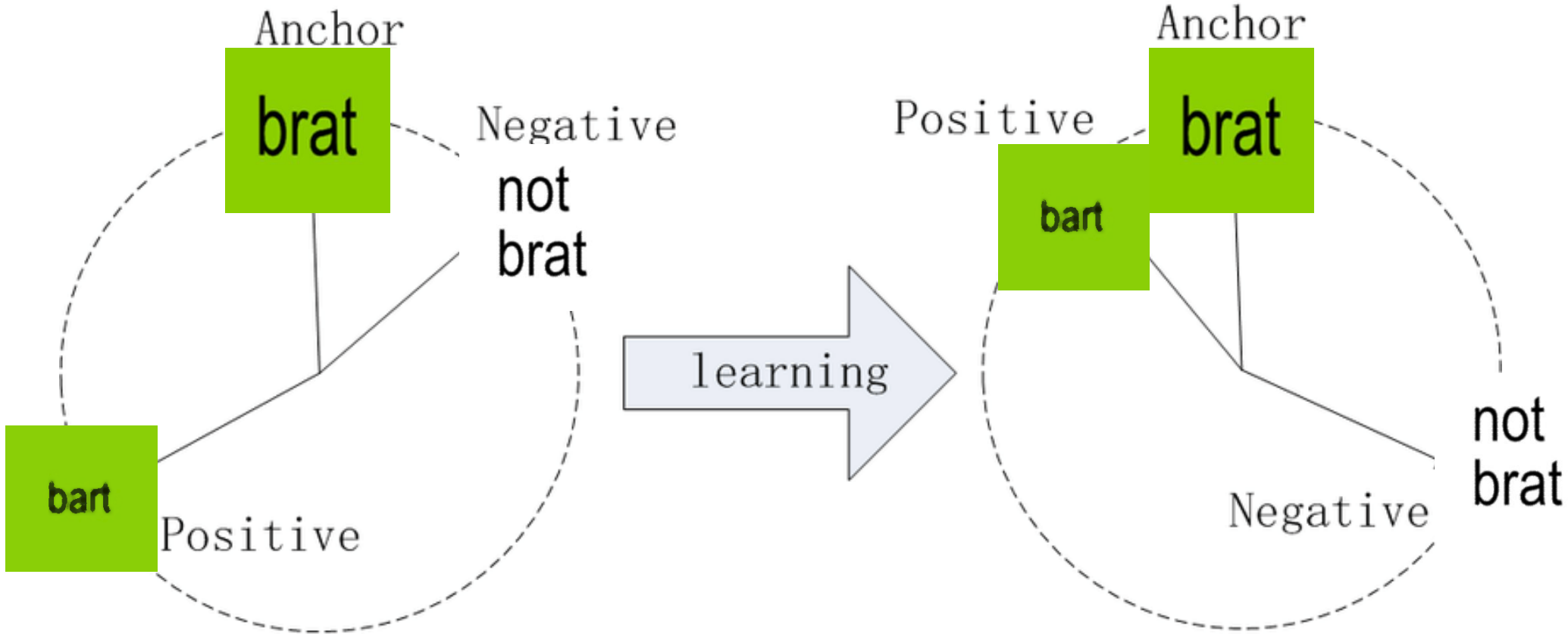


Triplet Loss

- Similar data points are closer together, and dissimilar ones are farther apart.
- In this case:
  - **Anchor**: Original Cover
  - **Positive**: Camera captured
  - **Negative:** A different cover



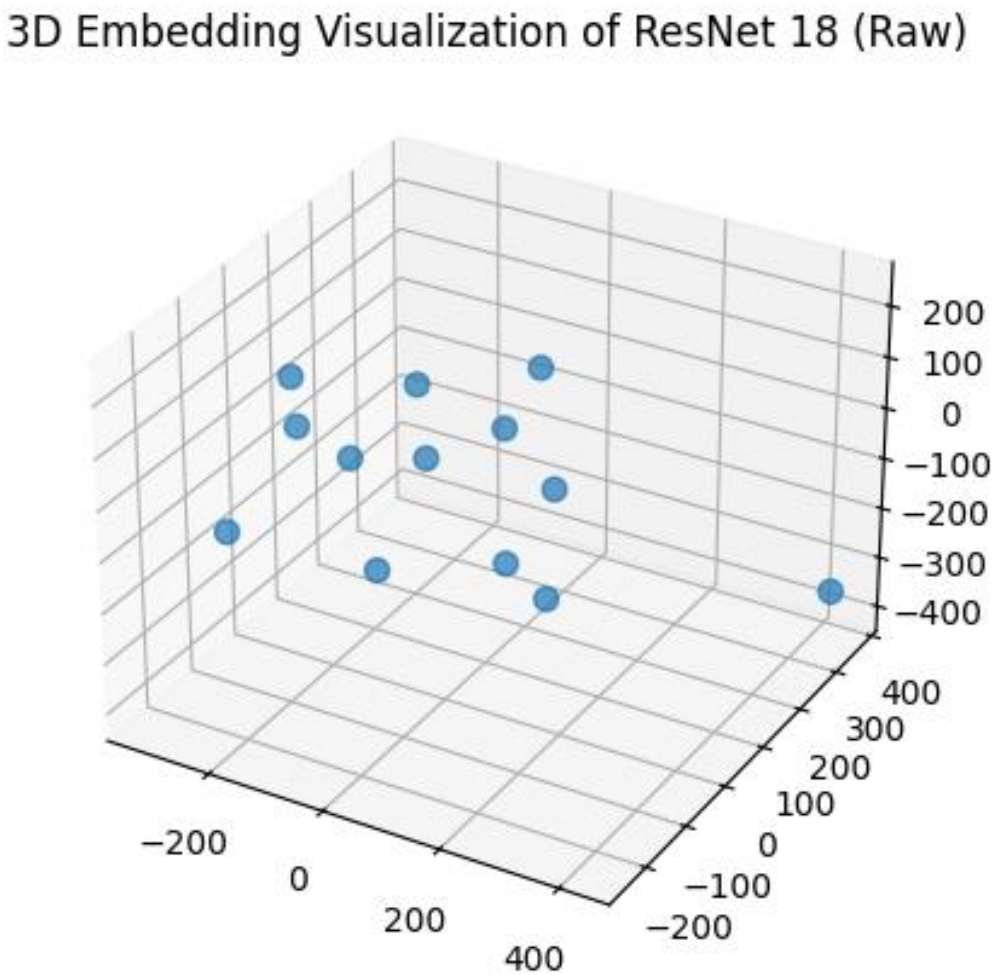
Triplet Loss

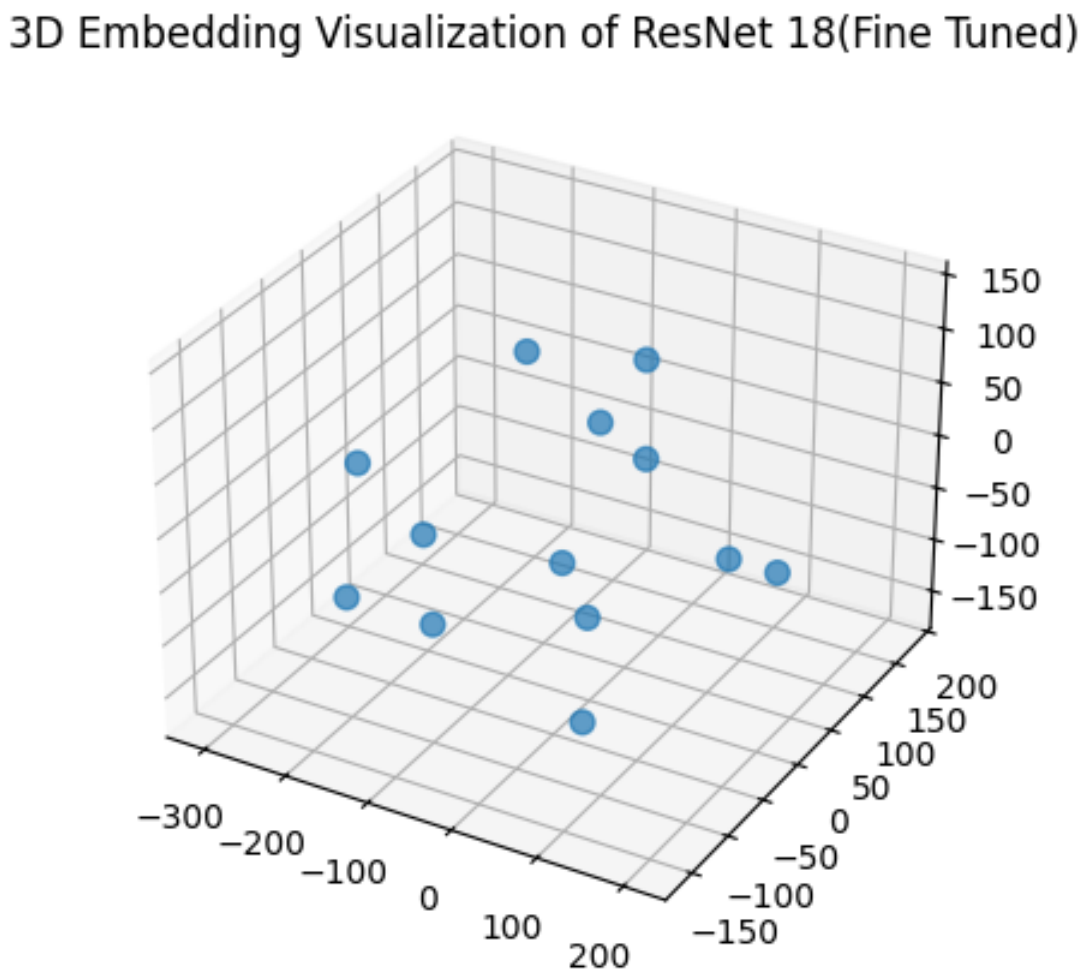


Transformation

- Approximate camera artifact.
  - Gaussian Blur is too expensive
- Colored Noise
- Random rotation
- Tilt (≤15 degree)
- ≤30% of area covered

Result


3D Embedding Visualization of ResNet 18 (Raw)
200
100
0
−100
−200
−300
−400
400
300
200
100
0
−100
−200
−200
0
200
400


3D Embedding Visualization of ResNet 18(Fine Tuned)
150
100
50
0
−50
−100
−150
200
150
100
50
0
−50
−100
−150
−300
−200
−100
0
100
200

Actual Results

- Mostly good
  - Struggles with certain album art types and lighting conditions
  - Non-deterministic results with multiple albums in frame
- Initial setup is still fairly involved for stations deploying

Acknowledgements

- Us four!
- Jeova
- WBOR
- RIP Dudley Coe
- Steve Jobs

wbor