EECS 442 Discussion: PS8

Dataset

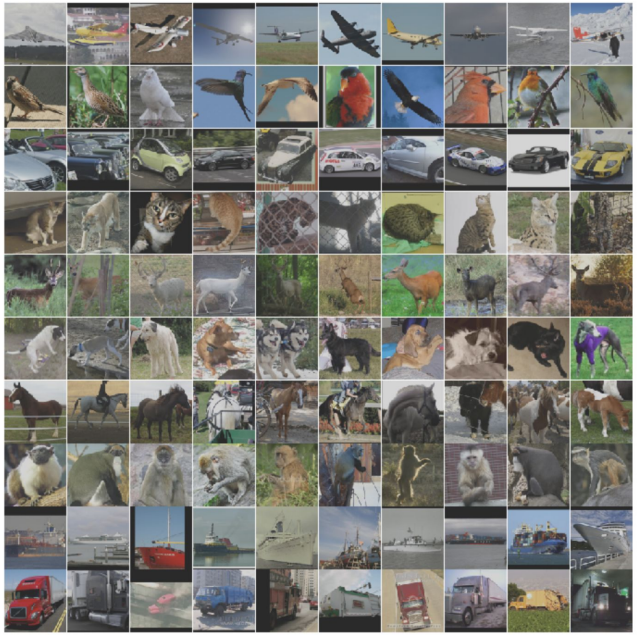
1. 10 classes: airplane, bird, car, cat, deer, dog, horse, monkey, ship, truck.

2. Images are 96x96 pixels, color.

3. 100,000 unlabeled images, 5,000 labeled training images, and 8,000 labeled test images

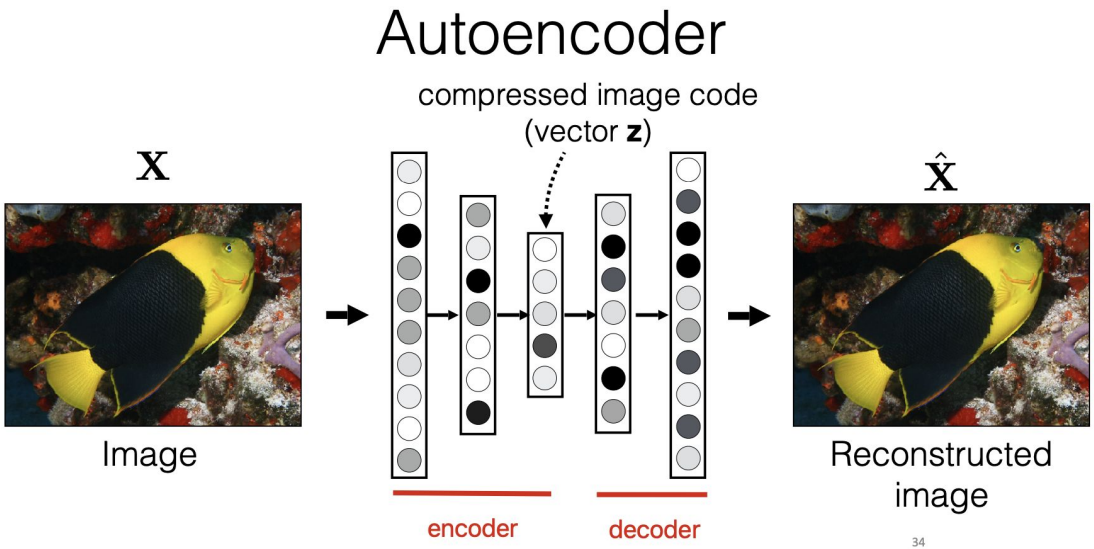
4. Use 100,000 unlabeled images to train autoencoder.

5. Use 5,000 labeled training images to train linear classifier, the input of the linear classifier is the representations learned from autoencoder.



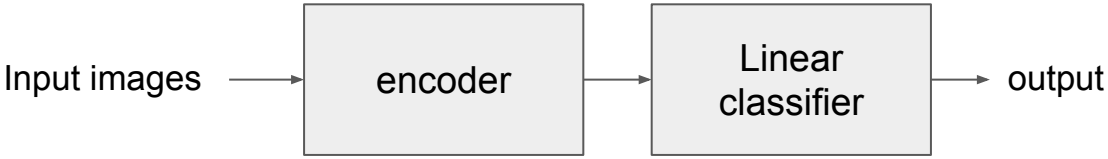
Autoencoder

We want to minimum the difference between original image and reconstructed image.
If the difference is small, the encoder has extracted a useful and concise representation of the original image.
We will use vector z as the input of linear classifier



vector z(output of encoder) is also the representation of the original image

Three methods



Method I: unsupervised pretraining + training linear classifier

Block the parameters optimization of the pretained encoder during the training of linear classifier.

1. Use 100,000 unlabeled images to train autoencoder

2. Apply the encoder of trained autoencoder on the 5,000 labeled training images

3. Use the output of step 2 as the input of linear classifier and train linear classifier.

4. Get classification accuracy on test images

Method II: supervised training

Directly train the encoder in the autoencoder together with the linear classifier.

1. Apply the encoder of autoencoder on the 5,000 labeled training images

2. Use the output of step 1 as the input of linear classifier, train linear classifier and the encoder together

3. Get classification accuracy on test images

Method III: random encoder + training linear classifier

Block the parameters optimization of the randomly initialized encoder during the training of linear classifier.

1. Apply the encoder of autoencoder on the 5,000 labeled training images

2. Use the output of step 1 as the input of linear classifier, train linear classifier

3. Get classification accuracy on test images

Lab color space

L: luminance(grayscale intensity)

A: green to red

B: blue to yellow

Use lab channels as the input of CMC Encoders

Contrastive Multiview Coding

Goal: get better representations than autoencoder

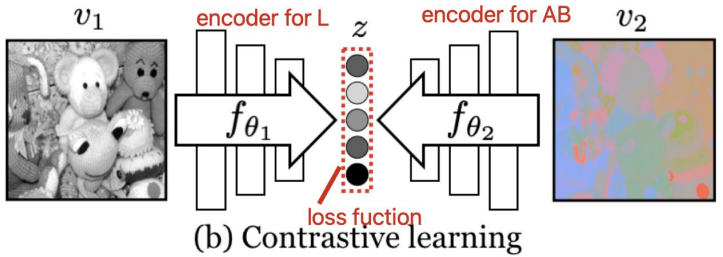
We will not minimize the difference between the original image and reconstructed image but to maximize mutual information in different views.

Hypothesis: a powerful representation is one that can contract view-invariant information

Views in CMC: views of an image that contain complementary information. (try to extract mutual and compact information)

The 2 views that we will use: L channel and combination of AB channels. (LAB together can form the colored image)

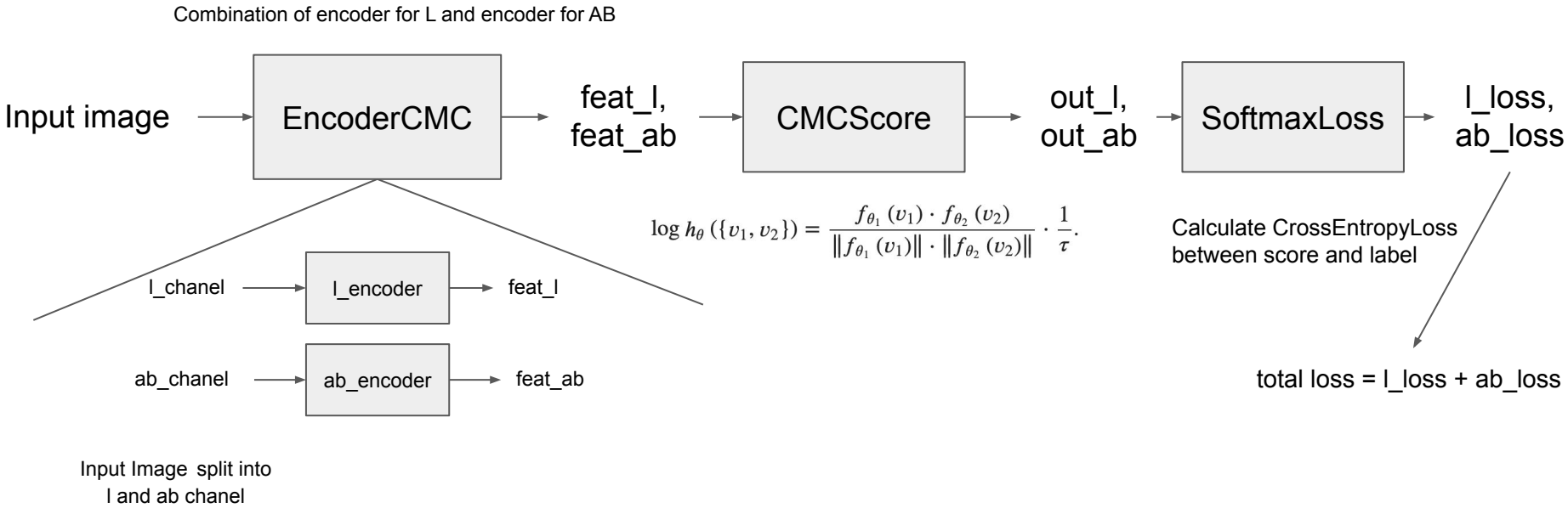
We will use two encoders to get the representations of L and AB.



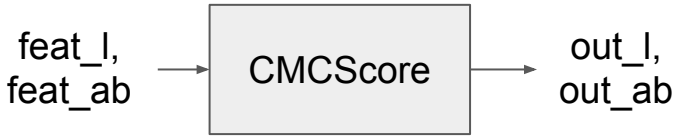
(b) Contrastive learning

V1: L channel of the image
V2: AB channels together of the image
$f\theta_1(\cdot)$ encoder for L
$f\theta_2(\cdot)$ encoder for AB

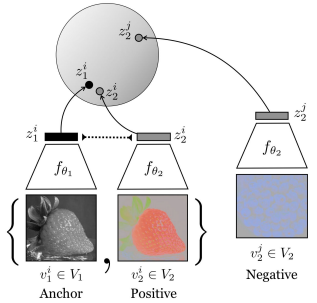
Structure of CMC



CMCScore



$$\log h_\theta\left(\{v_1, v_2\}\right) = \frac{f_{\theta_1}(v_1) \cdot f_{\theta_2}(v_2)}{\|f_{\theta_1}(v_1)\| \cdot \|f_{\theta_2}(v_2)\|} \cdot \frac{1}{\tau}.$$



Evaluate the similarity between $f\theta_1(v1)$ and $f\theta_2(v2)$, result has shape (batch size, K + 1). the 0-th element of the result is the score for positive pair.

We will consider the L and ab from the same image as positive pair and from K other images(random selected) as negative pairs.

We want $\log h_\theta\left(\{v_1, v_2\}\right)$ have high value for positive pairs and low value for negative pairs.

out_l is L as anchor, out_ab is AB as anchor

If v1 is anchor, $f\theta_1(v1)$ will have shape (batch size, feature dim), $f\theta_2(v2)$ will have shape (batch size, K + 1, feature dim).

Consider 1 image in the batch, K = 2,
out_l can be

0.06	Similarity score for positive pair
0.0	Similarity score for negative pairs
0.94	