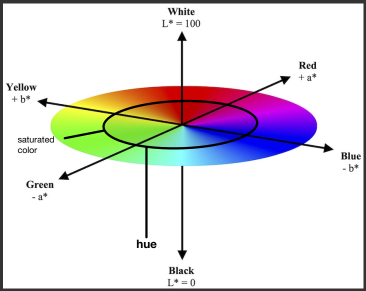
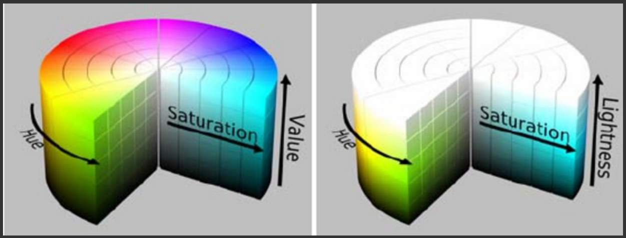
CIE L*a*b* color space :

L = luminance (0 to 100) | a* (green-red) (-128 to 127)
b* (blue-yellow) (-128 to 127)



Cylindrical view :
Think of croma (here a*, b*) defining a planar disc at each luminance level (L)

follow up : HSL (hue saturation lightness)
HSV (hue saturation value)



1 dimensional signal $\longrightarrow$ 3 dimensional signals

grayscale color channel

Grayscale Image : L channel

$X \in \mathbb{R}^{H \times W \times 1}$

input

f (high level abstraction / semantics)

Color Information : ab channels

$\hat{Y} \in \mathbb{R}^{H \times W \times 2}$

output

image synthesis problem

free supervisory signal

Concatenate (L, ab)

$(X, \hat{Y})$

refers to the idea that the data distribution has a single peak (mode)

Better Loss Function :

The L2 loss assumes a unimodal distribution of errors, meaning it is robust when the residuals are centered around zero with a single mode, following a Guassian (normal) distribution.

Colors in ab space (continuous)

Regression with L2 loss inadequate :

$$L_2(\hat{Y}, Y) = \frac{1}{2} \sum_{h,w} \left\| Y_{h,w} - \hat{Y}_{h,w} \right\|_2^2$$

If the underlying data or error distribution is multi-modal or has heavy tails, other robust loss functions might be more appropriate.

discrete bins of size $\boxed{10}$ (here)

Colors in ab space (discrete)

Use multinomial classification

$$L(\hat{Z}, Z) = -\frac{1}{H \cdot W} \sum_{h,w} \sum_{q} Z_{h,w,q} \log(\hat{Z}_{h,w,q})$$

Class rebalancing to encourage learning of rare colors :

$$L(\hat{Z}, Z) = -\frac{1}{H \cdot W} \sum_{h,w} v(Z_{h,w}) \sum_{q} Z_{h,w,q} \log(\hat{Z}_{h,w,q})$$

Previous works :

(i) non-parametric framework :

reference image (colors) $\longrightarrow \oplus \longrightarrow$ grayscale image $\rightarrow$ fail to generalize

(ii) parametric framework :

— L2 regression (may give sepia, or desaturated result)
(i) hand engineered features
(ii) deep networks

— Classification

Network Architecture :

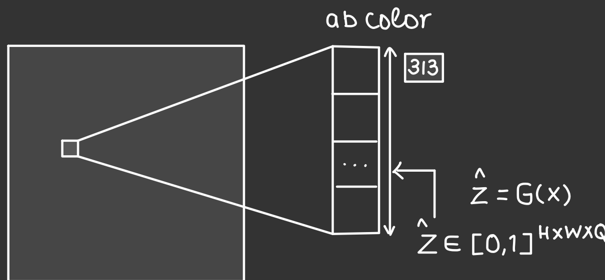
X = lightness (L channel)
$\hat{Y}$ = ab color (ab channel) $\xrightarrow{+L}$ L*a*b color image

$$\hat{Y} = F(X)$$

probability distribution — pixel level

VGG network :

Layers	dimensions
input X	224 X 224 X 1
conv1	224 X 224 X 64
conv2	112 X 112 X 128
conv3	56 X 56 X 256
conv4	28 X 28 X 512
conv5	14 X 14 X 512
fc1	1 X 1 X 4096
fc2	1 X 1 X 4096



VGG modified architecture:

Layers	dimensions
input X	
conv1	224 x 224 x 1
conv2	224 x 224 x 64
conv3	112 x 112 x 128
conv4	56 x 56 x 256
conv5	28 x 28 x 512
conv6	28 x 28 x 512
conv7	28 x 28 x 512
conv8	28 x 28 x 512
	256 x 256 x 56

Variation of standard convolutions in deep learning, where the convolutional kernel is expanded by inserting holes.
(larger receptive field without increasing the number of numbers or computation cost)

àtrous/dilated convolutions
(spatial resolution addition)

single point estimate:

$$\hat{y} = H(\hat{z})$$

interpolation between the mean and the mode, allows us to keep the vibrancy of the output colors while maintaining some spatial consistency.

Evaluation:

joint interaction b/w pixels & overall perceptual quality

	visual quality	representation learning
quantitative	- per pixel accuracy - perceptual realism - semantic interpretability	- task generalization - task & dataset generalization
qualitative	- low-level stimuli - legacy grayscale photos	- hidden activation units

Predicting Labels from Data:

Supervised training

$\boxed{\text{data } x}$ → $\boxed{\text{learned feature hierarchy}}$ → $\boxed{\text{label } y}$

imageNet images | imageNet labels

Predicting Data from Data:

Supervised learning

$\boxed{x_0 \mid x_1}$ → learned feature hierarchy → $\boxed{\text{label } y}$

imageNet images | imageNet labels

Unsupervised learning (self-supervised training)

$\boxed{x_0}$ → learned feature hierarchy → $\boxed{x_1}$