ArtTalk: Labeling Images with Thematic and Emotional Content

[Yifan Wang]

Background

A picture is worth a thousand words. Through history of mankind, artists waves their brush for two purposes: depict the object or convey a mood. Therefore, I want to teach computer to view artistic paintings like human beings by telling the them and emotion in the art work.

This is critical for advertising and broadcasting business since this can automate the procedure for picking the right picture for given events.

Theme:
- Lots of work on obj. detection in photo but few on painting
- Transfer learning needs tuning

Emotion:
- Some facial emotion detection but few on painting
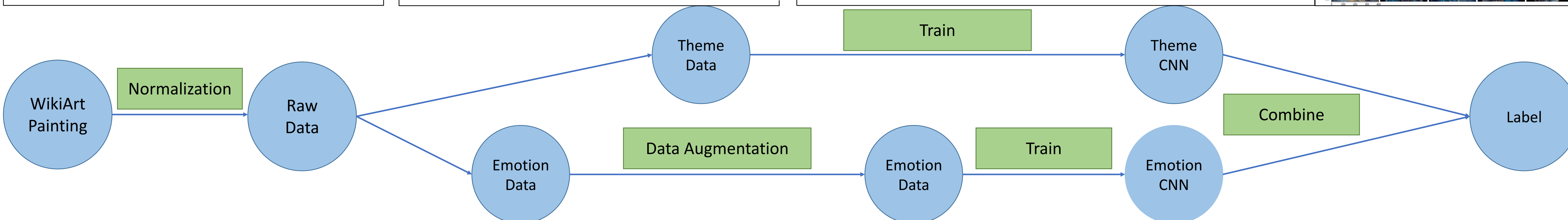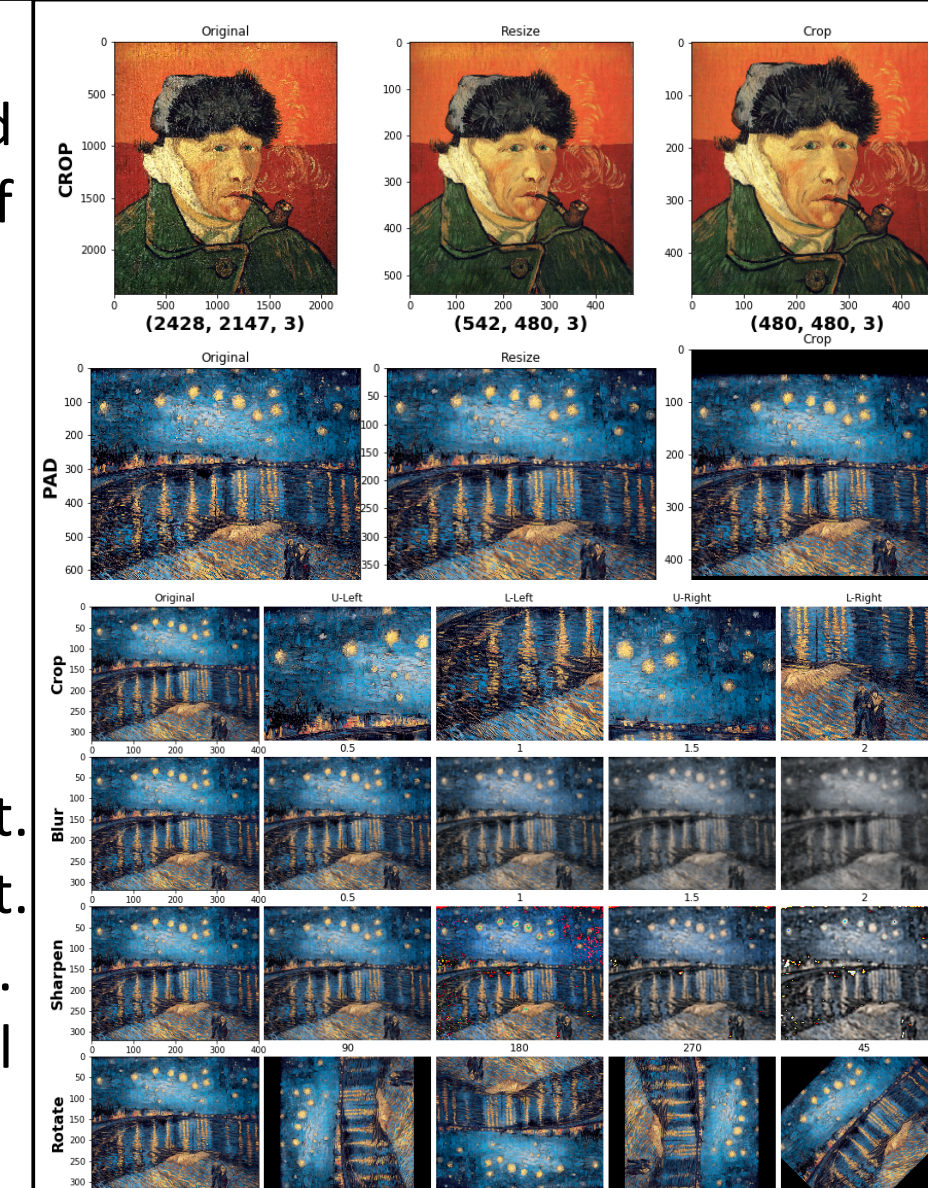- Existing ways get less than 50% correct

Dataset

35750 Art Paintings are downloaded from WikiArt and they are used for both tasks.
- Theme Detection

WikiArt pictures have been labeled with theme. I choose the themes with more than 150 pictures. 33473 pictures are chosen covering 41 themes.
- Emotion Detection

2000 pictures are manually labeled with 11 feelings: **neutral**, happy, love, joy, sad, disgust, fear, surprise, lust, anger and envy. They are then augmented to 34000 training images.

	Testing	Training	Fold	Augmented
Theme	1473	32000	8	No
Emotion	2000	32000	8	Yes

Algorithm

- Data Normalization: To Render all the pictures into the same size. First I reshape all image to the same width. Then I either crop or pad along height to render the images into the same shape. Size of 224X224 is chosen finally as a consequence of limited memory.
- Data Augmentation: To provide more data for emotion learning.
  - ✓ Crop: upper left, upper right, lower left, and lower right.
  - ✓ Blur: different blur strength from 0.5 to 2.
  - ✓ Sharpen: different sharpen strength from 0.5 to 2.
  - ✓ Rotate: 45°, 90°, 180°, and 270°. Unfitted parts are cut.
- CNN Structure:
  - ✓ VGG 16, with additional average pooling layers for output.
  - ✓ ResNet, with additional average pooling layers for output.
  - ✓ [Gudi et al], [Xu, et al], [You et al], [KARAYEV ET AL]. These are the structures proposed and tested for facial emotion detection. I am tuning the ensemble of them.
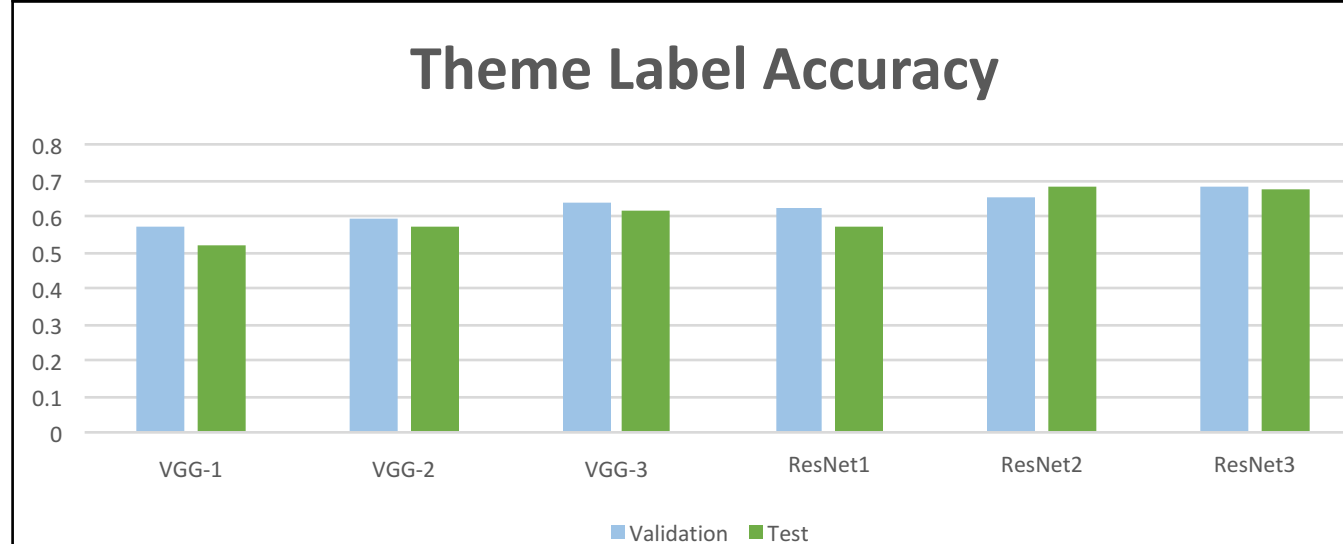

Flow Diagram

WikiArt Painting → (Normalization) → Raw Data → Theme Data → (Train) → Theme CNN → (Combine) → Label

Raw Data → Emotion Data → (Data Augmentation) → Emotion Data → (Train) → Emotion CNN → (Combine) → Label

Problem Statement

- **Problem Definition**

The goal of ArtTalk is to tag a picture with theme label and emotion label. Theme labeling differs from object detection since abstract themes like Christianity are included. Emotion labels cover 10 major human feelings.

- **Approaches**

Two CNNs are trained for either task. Then we combine the output of two CNNs to form the labels of the painting.

During training, transfer learning and ensemble are used for higher accuracy. I have also tried novel network structures.

- **Evaluation**

The results will be evaluated on part of raw data, which is used as test sets. Percentage of correct labeling will be reported separately as accuracy rate for either model.
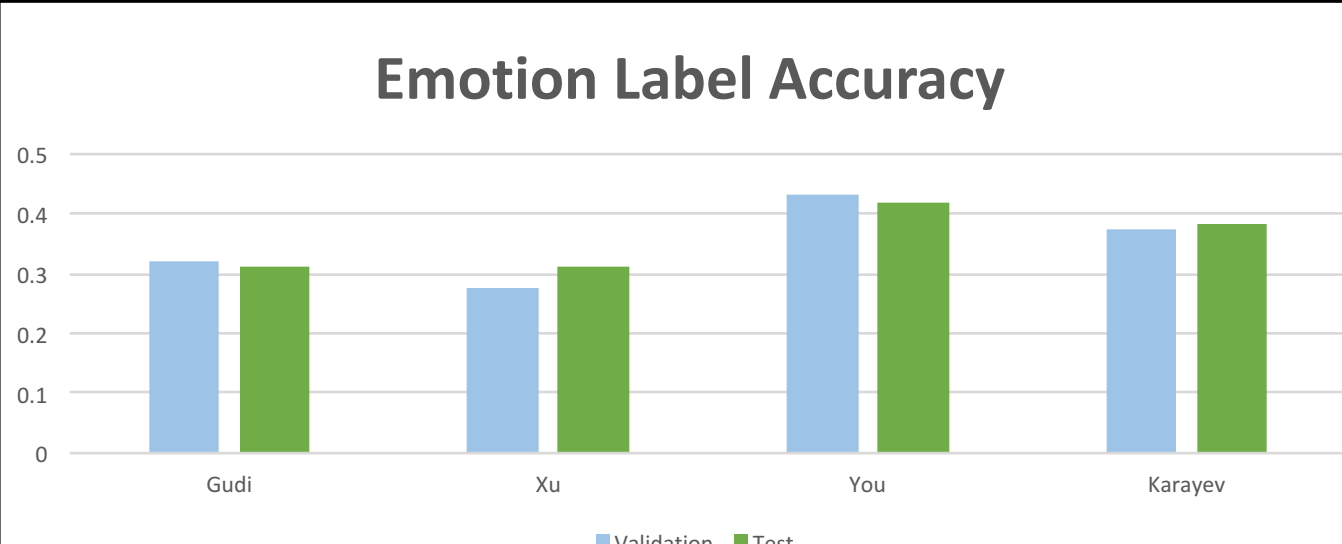
Evaluation

Theme Label Accuracy

I have tried six different structures:
- VGG-1: VGG16 + avg pooling
- VGG-2: VGG16 + conv-layer + avg pooling
- VGG-3: VGG16 + two conv-layers + avg pooling
- ResNet-1: ResNet + avg pooling
- ResNet-2: ResNet + conv-layer + avg pooling
- ResNet-3: ResNet+ 2 conv-layers + avg pooling. This provides the best test accuracy of 67.9%.

I am working on better tuning ResNet-3 structure to deliver a higher accuracy.

Emotion Label Accuracy

I have tried four existing network structures for facial expression emotion recognition and the best accuracy I get is around 43%.

I have encountered the following issues:
- Difficult to differentiate subtle feelings within the same sentiment like happy and love.
- Training data do not follow uniform distribution and top three emotions contain more than half of training images. This make it difficult to learn emotions like lust, with few data.

Future Work

- Tune ResNet-3 for higher accuracy.
  - ✓ Ensemble of models will be used
  - ✓ Aim at an accuracy higher than 80%
- Propose new structures for emotion detection
  - ✓ Consider more features like LBP.
  - ✓ Use more complex structures and deeper network.
  - ✓ Try better optimization algorithm.
- Data Visualization:
  - ✓ To tell which parts of a picture affect our mood most.
  - ✓ To tell which color / illumination affect our mood.
  - ✓ To tell which object on the painting affect the choice of theme, especially for abstract themes.
  - ✓ Use saliency map for the task.