Digital Image Processing Project

Sky is Not The Limit : Semantic aware color transfer

Anjali Shenoy (201401114), Saumya Rawat(201401110)

Introduction

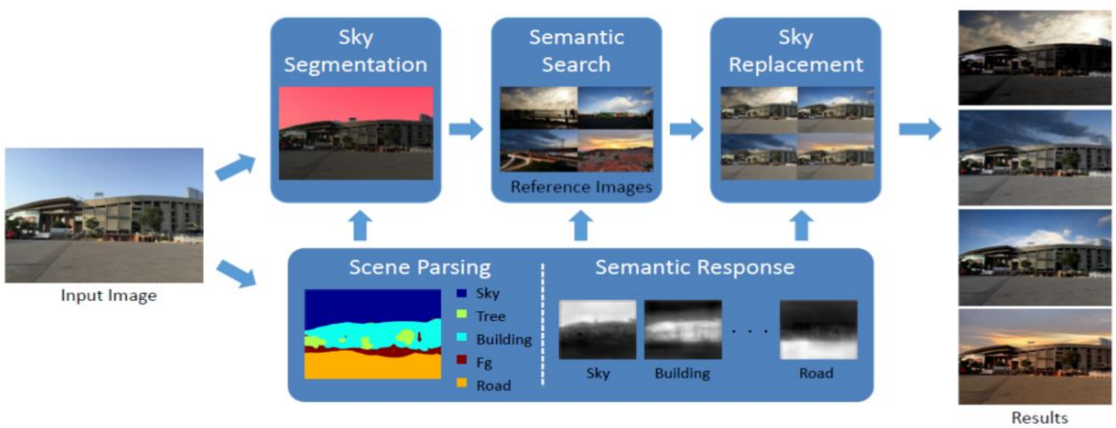
Skies are one of the most common backgrounds in photos. However, we have no control over the weather or lighting conditions at the moment of photography. As a result, numerous interesting and valuable photos have uninteresting or poorly exposed sky regions. Professional photographers fix this problem using sophisticated tools by manually delineating the sky regions precisely, testing different skies for compatibility, and finally adjusting the foreground to match the new composited sky. This requires time and expertise that is beyond the abilities of novice users.

Hence in this project we have implemented a fully automatic sky replacement tool that can take an input image and generate a set of realistically edited photos with interesting skies and different styles.

Problem Statement

To replace sky of an input image I with that in a reference image R while also doing a label guided color transfer of the luminance and chromaticity of the reference image into the input image.

Algorithmic Overview



Pipeline

To achieve this goal of semantic aware sky replacement, we address three challenging tasks:

- Sky segmentation
- Retrieve candidate skies
- Sky replacement
- Creating realistic transfer according to the new skies

Sky Segmentation

Sky appearances vary widely among images, and without an understanding of scene layout, it can be indistinguishable from non-sky regions (e.g., reflections in water).

We use a deep Fully Convolutional Neural Network (FCN) to parse an input image and generate a dense pixel-wise prediction of semantic labels such as sky, tree, building, mountain and water. By understanding the global layout of the scene, the proposed algorithm can robustly localize the sky regions in spite of different colors, shapes, sizes and attributes. The paper proposes a graph cut based energy minimization next to refine the 95% accurate segmentation obtained by the fcn output.

$$E(X) = \lambda_1 \sum_i U_c(x_i) + \lambda_2 \sum_i U_t(x_i) + \lambda_3 \sum_i U_f(x_i) + \lambda_4 \sum_{(i,j)\in\mathcal{E}} V(x_i, x_j),$$

Where the values $\lambda_{\{1,2,3\}}$ are weights set to 1 and $\lambda_{\{4\}}$ is 100. U_c and U_t are color and texture unary potentials for the sky and non-sky label costs (obtained from the learned online classifier), and U_f accounts for the FCN output. In addition, V is the pairwise potential for smoothness in a set E of adjacent pixels.

As they have already provided the output of this step we are using the parsed masks directly.

(a) Input image (b) FCN result (c) Fine segmentation (d) Input image (e) FCN result (f) Fine segmentation

Sky Search

According to the model suggested by the paper, the descriptors for dataset of 414 images are computed based on the response map obtained from the FCN output which is probabilities for each pixel (i,j) of belonging to either of the 14 labels. Since the color transfer depends on the labels of parts of the image it is important that the retrieved options for reference images are such that the semantic layout and content match the input image. To ensure this the descriptor is constructed using a spatial pooling method wherein the image is divided into 9 grids. For each grid, $h^j_{i...m} = 1/m * \Sigma_i f^j_i$, is calculated and the final descriptor $H = [h^1, h^2 .. h^{14}$, global histogram] for each image is constructed. This is used for pairwise matching with the input and top 4 candidates are retrieved. Another constraint on the retrieval is the aspect ratio and resolution of the pair, reference and input i.e Q value defined as $\min(P_I, P_R) / \max(P_I, P_R)$ is above 0.5, where P is both aspect ratio (width/height) and resolution (width*height) to ensure the two images are similar in terms of areas of sky.

Sky Replacement

Once the reference sky image is selected out of the retrieved images, the largest rectangular convex hull enclosing the sky region in the input image is computed. Then the maximum rectangle enclosed in the reference sky region is computed, this rectangular area is then resized to match the convex hull to obtain a rectangle, say A. The sky pixels in the input image are given values from A. This image then serves as input for the next phase in the pipeline.

Semantic aware transfer

Transferring color statistics from one image to another is a common technique in image editing. Existing approaches usually perform transfer over the entire region without taking visual semantics into account [Reinhard et al. 2001; Tao et al. 2009] and generate less

realistic results when the image content are not well matched.

Suppose the total number of semantic labels existing in the scene parsing map of the input image is n r , we formulate the transfer process for each pixel x as:

$$T(x) = \sum_{n=1}^{n_r} W_n(x) \cdot T_n(x),$$

where $W_n(x)$ is the likelihood value in the normalized FCN response map on pixel x for semantic category n, and $\sum_n W_n(x) = 1$. For each semantic label n in the scene parsing results of the input image, we compute a category-specific color/luminance transfer function T_n.

T_n <u>when label n is present in both original image and reference image:</u>

I_n = region associated with Input image having label n

R_n = region associated with reference image having label n

- Compute the color difference between I^{sky} region and R^{sky} region in LAB space by taking the absolute difference of the means of their color. Pass this through a limiting function such as $tanh$ so that difference is always within a range of 0

$$\beta = tanh(|c(I^{sky}) - c(R^{sky})|)$$

- Use this β as a regularisation constant when computing the desired luminance of input image using I_n and R_n.

$$\overline{L} = L(I_n) + \beta(L(R_n) - L(I_n))$$

$L(I_n)$=Mean of luminance of Input image for category n

$L(R_n)$=Mean of luminance of Reference image for category n

- For chrominance transfer we model the chrominance distribution of an image using a multivariate Gaussian, and find a transfer function that maps the Gaussian statistics $\eta(\mu_R, \Sigma_R)$ of Reference image to $\eta(\mu_I, \Sigma_I)$ to Input image.

$$c_{output}(x) \;=\; T(c_I(x) - \mu_I) + \mu_R$$

$$T = \Sigma_I^{-1/2}(\Sigma_I^{1/2}\Sigma_R\Sigma_I^{1/2})^{1/2}\Sigma_I^{-1/2}$$

where T is a linear transformation that maps chrominance between the images and $c(x)$ is the chrominance at pixel x.

To avoid color artifacts due to low chrominance we clip the diagonal elements of Σ_I as

$$\Sigma_I^{'} \;=\; max(\Sigma_I,\; \lambda_r I)$$

where λ_r = 7.5 is a regularisation constant

Transfer function when label in input image is not in reference image:

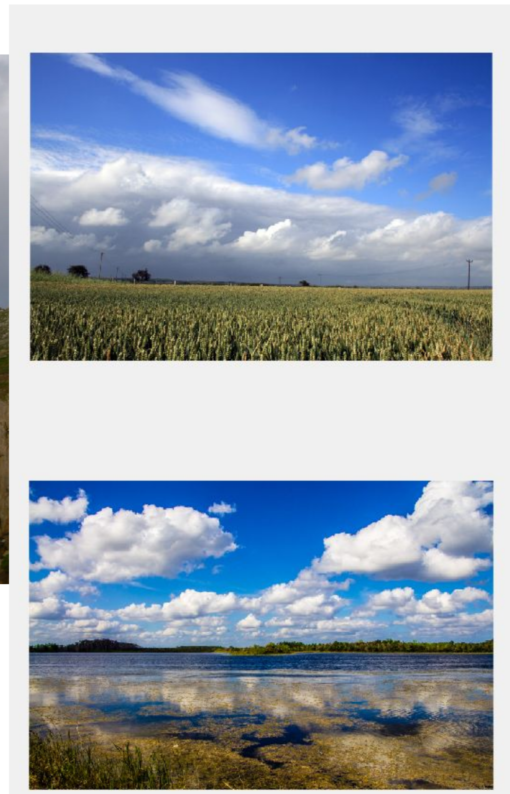
In the second case, when there is no matched region found in the reference image for label n, we resort to the finding the regions neighbouring the label n in Input image and try and see if this exists in reference too. If it exists we transfer luminance and chromaticity from this to our input image, otherwise we let them be as it is.

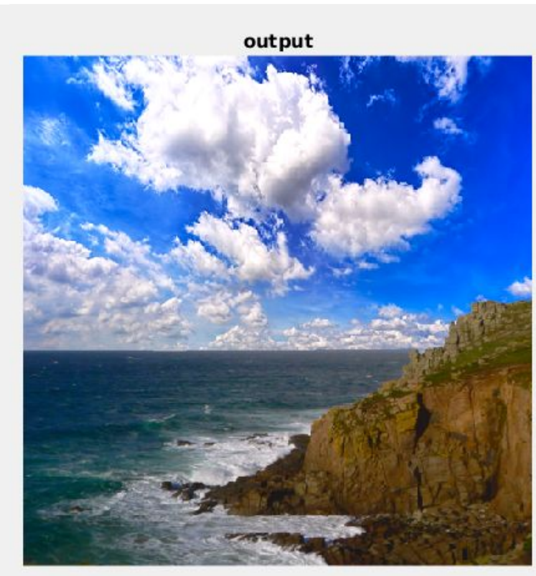
The idea behind this was that if we don’t get the label in the reference image, the label in the input image should have similar Luminance and chromaticity as its neighbour. Hence we transfer this from reference image to input image.

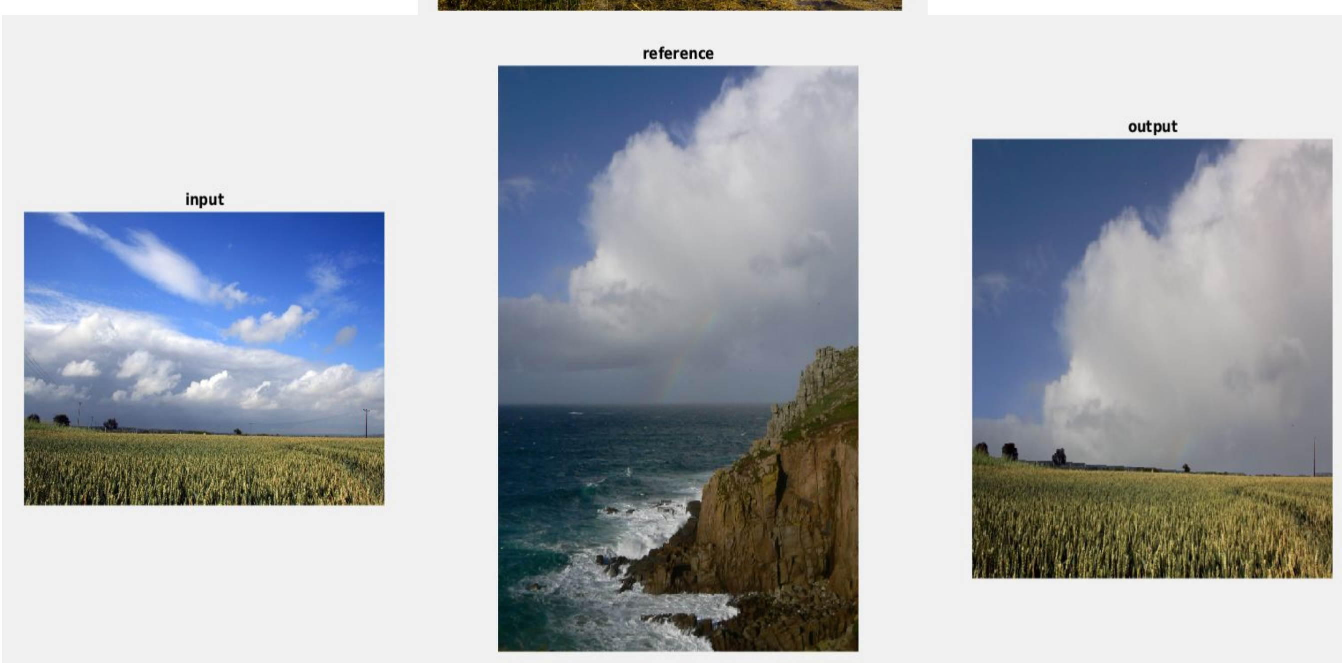
Results

1) Basic Pipeline of the model

Input Image : Retrieved Images : Output:

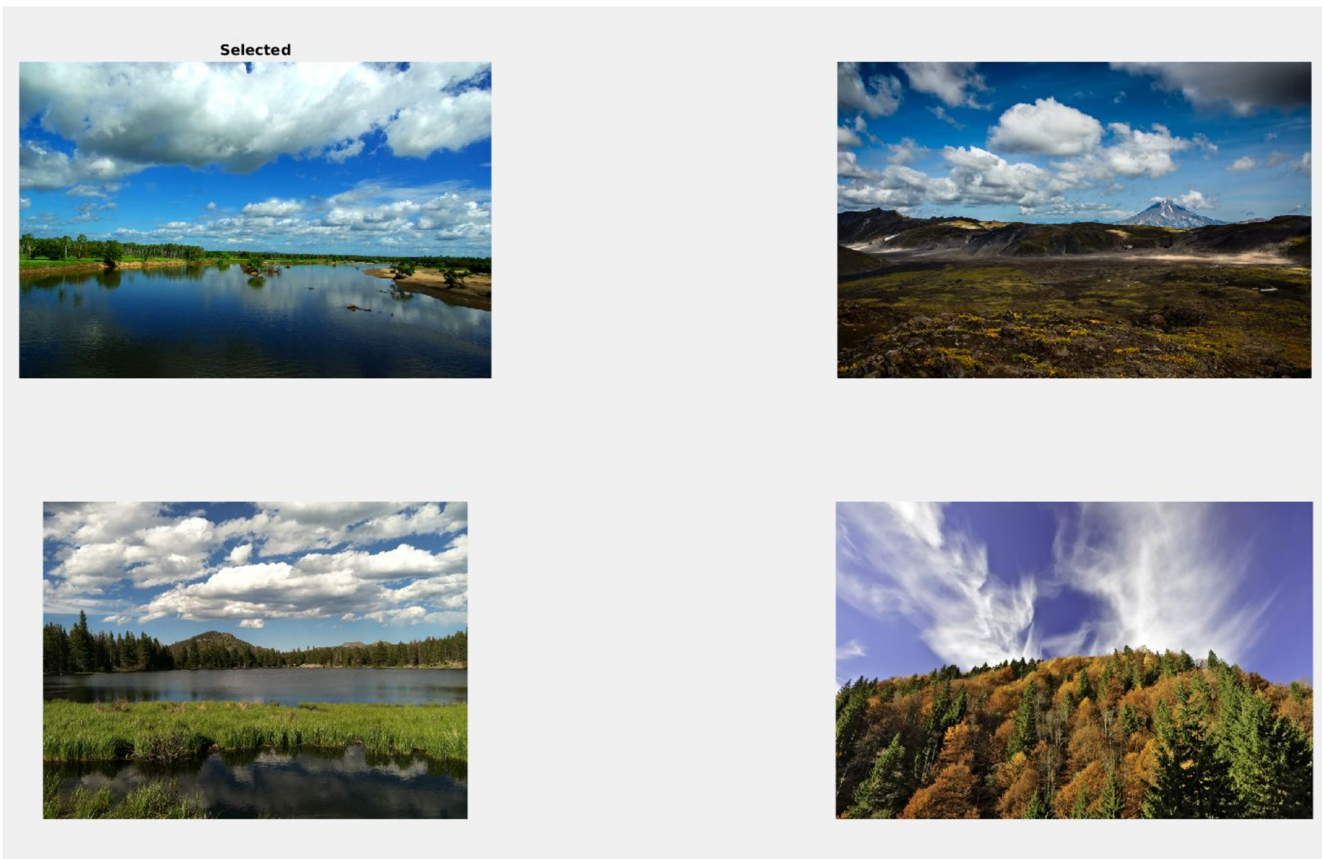






2) Retrieval of skies

Input Image :

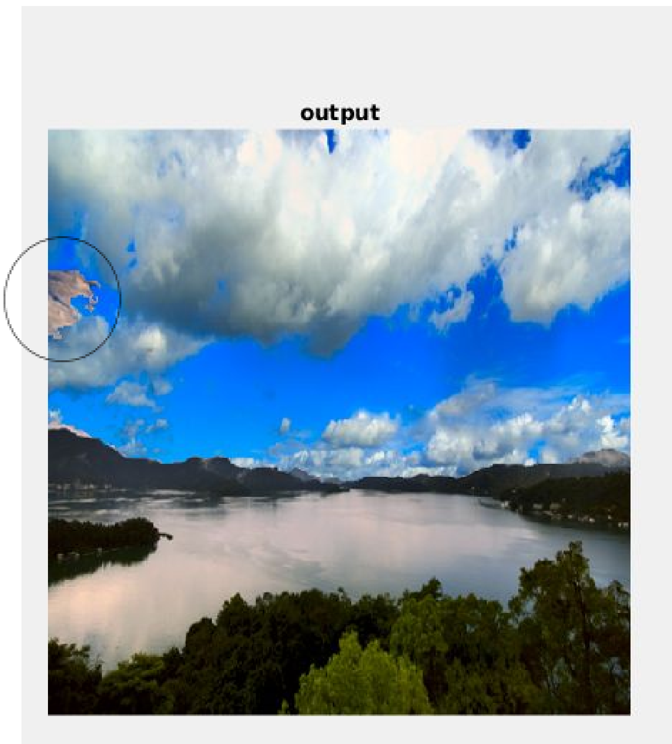


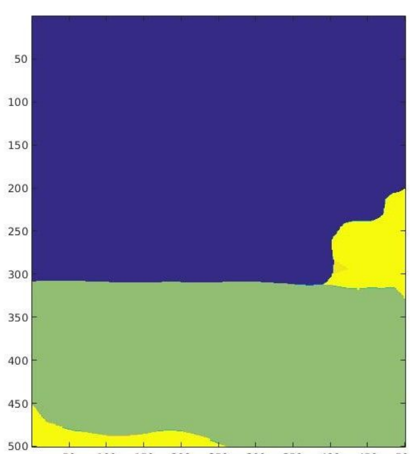
3) 4 Different outputs

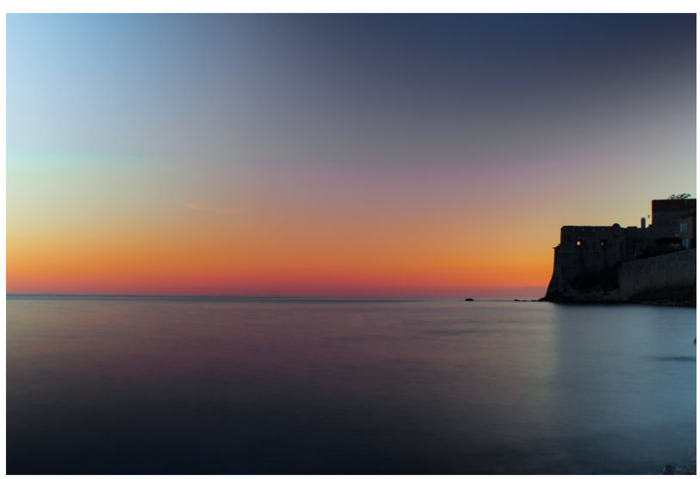
4) Miscellaneous

Challenges

- FCN output gave 95% accuracies, so there were some cases where the labelings were mismatched. Eg. labeling a sky reflection on the sea as sky. This error poses a problem in both sky replacement as well as colour transfer.







- Computing the descriptors and FCN response map for 414 images was computationally heavy. Each pixel of each image had a 1xN dim. vector attached to it to give the probability of that pixel belonging to any of the N categories. Writing optimal code, Importing, using, and storing this was quite challenging.
- The dataset used by the paper is not diverse enough to correctly support the retrieval of images. It consists *majorly* of images with mountains and lakes and hence retrieval for images with say, buildings becomes an outlier case and is unable to retrieve semantically similar images in that particular case. The images are not equally divided for all labels.
- This kind of color transfer is less effective for images with strong directional lighting or high-level cues like shadow directions and reflections.

Conclusion

In this project, we implemented an automatic colour transfer method that is based on semantic information for creating images with stylized sky backgrounds. For a given input image, several skies are retrieved from the database and the selected sky is applied to the image with aesthetically realistic colour properties. For future work, the scene parsing fcn code slows down the process and becomes a bottleneck for computation time. We propose clustering the images based on their non-sky properties with a feature vector consisting of properties that have been shown to be highly correlated between sky and non-sky regions. These clusters being similar in properties can directly choose a reference sky and perform replacement, hence removing the need to use labels and perform colour transfer.

References

1. http://www.eecs.harvard.edu/~kalyans/research/skyreplace/SkyReplacement_SIG16.pdf
2. Automatic Content-Aware Color and Tone Stylization [Lee et. al]
3. Understanding and Improving the Realism of Image Composites [Xue et al]