1 Technical Details

1.1 Loss function formulations

Foreground Colorization. Let $x$ be an input object instance sketch image, $y$ the corresponding ground truth image, and $s$ the paired input natural language expression. The GAN objective function is expressed as:

$$L_{GAN}(D, G) = \mathbb{E}_{y \sim P_{image}}[\log D(y)] + \mathbb{E}_{x \sim P_{sketch}, s \sim P_{text}}[\log(1 - D(G(x, s)))],$$

and $L_{GAN}(G)$ uses the second term in this equation.

Let $c$ be a class label output by the discriminator $D$. The auxiliary classification loss $L_{ac}(D)$ for $D$ is defined as the log-likelihood between the predicted and the ground-truth labels:

$$L_{ac}(D) = \mathbb{E} \log P(C = c|y).$$

The auxiliary classification loss $L_{ac}(G)$ for generator $G$ is defined in the same form as $L_{ac}(G) = L_{ac}(D)$ with the discriminator fixed but the image to be classified as a synthesized one.

The supervision loss $L_{sup}(G)$ and the complete loss functions $L(D)$ and $L(G)$ for foreground colorization can be found in Equation 2, 3, and 4 of the main paper.

Background Colorization. Given the input image $x$ with the partially or completely colorized foreground objects, the ground-truth color image $y$, and the language description $s$, the generator $G$ produces the synthesized image with the colorized background $G(x, s)$. The cGAN objective function is expressed as:

$$L_{cGAN}(D, G) = \mathbb{E}_{x \sim P_{fg}, y \sim P_{image}}[\log D(x, y)] + \mathbb{E}_{x \sim P_{fg}, s \sim P_{text}}[\log(1 - D(x, G(x, s)))],$$

and the objective of the generator $L_{cGAN}(G)$ is to minimize the second term.

Given the category size $C$, the segmentation mask prediction $\hat{R} \in \mathbb{R}^{W \times H \times C}$, and the ground truth segmentation mask $R$, the segmentation loss $L_{seg}(G)$ is expressed in a cross-entropy manner:

$$L_{seg}(G) = -\frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} \sum_{k=1}^{C} \left( \hat{R}_{kj}^i \log(R_{kj}^i) \right).$$

The supervision loss $L_{L1-sup}(G)$ and the complete loss functions $L(D)$ and $L(G)$ for background colorization can be found in Equation 5, 6, and 7 of the main paper.

*Both authors contributed equally to the paper.
†Corresponding author: mcsgcy@mail.sysu.edu.cn
‡This project was started before this author joined Google.
1.2 Implementation Details

**Instance Matching Experiments.** The maximum training iteration was 100K and the batch size was set to 1. The initial learning rate was set to 0.00025 and Adam [2] was used as the optimizer. We resized the scene sketch images and the corresponding ground-truth masks to $768 \times 768$. The iteration numbers of LSTM and mLSTM were both set to 15. The cell sizes of LSTM and mLSTM were respectively set to 1,000 and 500. The Deeplab-v2 model [1] was trained on the SketchyScene dataset [4].

**Foreground Instance Colorization Experiments.** We set the maximum training iteration to 100K and used a mini-batch size of 2. We employed Adam [2] as the optimizer and set the initial learning rate of generator to 0.0002 and that of discriminator to 0.0001. The iteration numbers of LSTM and mLSTM were both 15 and their cell sizes were both set as 512. We set $\lambda_1 = 1$ and $\lambda_2 = 100$ in Equations 3 and 4 in the main paper.

**Background Colorization Experiments.** We trained 100K iterations using a mini-batch size of 1. Adam optimizer was used and the initial learning rate for both generator and discriminator was set to 0.0002. The iteration numbers of LSTM and mLSTM were both 9 and their cell sizes were both 1024. We set both $\lambda_1$ and $\lambda_2$ at 100 in Equation 7 in main paper.
2 Data Collection Details

2.1 Data Collection for Instance Matching

To train the instance matching network, we require triplet samples of scene sketches, text descriptions, and instance mask(s) as shown in Figure 6 in the main paper. Since collecting such a kind of data through manual annotation requires enormous crowdsourcing efforts, we designed and implemented a fully automatic rule-based algorithm to generate the paired data, based on some insights we learnt from the SketchyScene data [4] and the human cognition as below:

- The 24 selected categories (as shown in Table 1 of the main paper) can be divided into several higher-level groups based on their characteristics, as shown in Table 1.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distant objects</td>
<td>sun, moon, cloud, star</td>
</tr>
<tr>
<td>Still objects</td>
<td>house, bus, truck, car, bench, tree, road, grass</td>
</tr>
<tr>
<td>Animated objects</td>
<td>bird, butterfly, cat, chicken, cow, dog, duck, horse, people, pig, rabbit, sheep</td>
</tr>
</tbody>
</table>

- Humans tend to describe the adjacent objects with the same category using a single expression, e.g. “the two trees on the left are green”.
- Humans tend to describe distant objects without other reference objects or spatial information, e.g. “the clouds are light blue” / “all the stars in the sky are red”.
- For still objects, humans tend to describe them without other reference objects but with optional spatial information, e.g. “the left house is red with black roof” / “all the grass are dark green” / “the road is black”.
- For animated objects, humans tend to describe them with still objects as reference along with optional spatial information, e.g. “the person near the left car is in blue” / “the second chicken on the right is yellow” / “the dog has brown body”.

Based on these insights, we designed a fully automatic rule-based algorithm, which is summarized in Algorithm 1. In this algorithm, we obtained the language expression describing the location of an instance, e.g. “the tree in the middle” / “the bus”, as well as its binary mask as shown in Figure 6 of the main paper. However, in practice, the instructions that users assign to the system specify not only the instance of their interest, but also their colorization goal, such as “the tree in the middle is green”. To construct such a fully automatic model which still works well on distinguishing specified target(s) based on an expression even with extra colorization information, we turned to augmenting the location-only expression with random colorization descriptions. For example, after obtaining “the bus”, we randomly selected a colorization description designed for bus, e.g. “has orange body and blue windows”, from the FOREGROUND dataset, thus producing “the bus has orange body and blue windows” finally. Note that data collection for the instance matching task was automatically completed without any manual annotation.
Algorithm 1: Instance Matching Data Generation

**Input:** bboxes $B : [B_1, B_2, \ldots, B_n]$, class_labels $L : [L_1, L_2, \ldots, L_n]$, masks $M : [M_1, M_2, \ldots, M_n]$

**Output:** a set of $O$ \{caption $T$: its corresponding masks $[M_p, M_q, \ldots]$\}

1. for $B, L \in B, L$ do
   2. raw_items = RegisterItem($B, L$)
   3. $distant\_items = SelectDistantItems$ (raw_items)
   4. $O\_dist \{T\_dist : [M_p, M_q, \ldots]\} = GetTextAndMasksByItemNumber(distant\_items, M)$
   5. $near\_items = SelectNeartItems$ (raw_items)
   6. $still\_items, animated\_items = SplitItems$ (near_items)
   7. $Function\ GroupingAdjacentItems(items)$:
      8. for item $\in$ items do
         9. recursively look for another item $t \in$ items
         10. if IsSameCategory (item $t$, item) & NotGrouped (item $t$) & IsAdjacent (item $t$, item) then
            11. item$\_groups = MakeItemGroups$ (item $t$, item)
            12. item$\_groups$$_{map} = \{item\_groups.category: item\_groups\}$
            13. return item$\_groups$$_{map}$;
      14. $still\_groups = GroupingAdjacentItems$ (still_items)
      15. $animated\_groups = GroupingAdjacentItems$ (animated_items)
   16. $Function\ SetPositionOfItemsWithinGroup(group)$:
      17. SortByHorizontalPos (group)
      18. pos_distribution = FindPosDistribution (group)
      19. for item $\in$ group do
         20. item.SetPosition (pos_distribution)
      21. return;
   22. $Function\ FindReference$self$\_groups, ref\_groups$:
      23. SortByHorizontalPos (self$\_groups$)
      24. for s$\_group \in$ self$\_groups$ do
         25. if IsEmpty (ref$\_groups$) then
            26. ref = FindClosestRefWithinSelfGroup (self$\_groups$)
            27. if IsNotEmpty (ref$\_groups$) then
               28. ref = FindClosestRefWithinRefGroup (ref$\_groups$)
               29. s$\_group$.SetReference (ref)
               30. SetPositionOfItemsWithinGroup (s$\_group$)
         31. return;
   32. $FindReference(still\_groups, [] )$
   33. $FindReference(animated\_groups, still\_groups)$
   34. $O\_near \{T\_near : [M_p, M_q, \ldots]\} = GetTextAndMasksByRefAndPos(still\_groups + animated\_groups, M)$
   35. $O = O\_dist + O\_near$
2.2 Data Collection for Foreground Instance Colorization

The foreground instance colorization task requires triples of cartoon image, edge map (sketch), language description, as shown in Figure 7 of the main paper. The detailed procedure of data collection for this task is described below:

1. We first crawled cartoon instance images, covering 24 object categories, from the Internet and then leveraged X-DoG [3] to extract an edge map as the corresponding sketch for each image. All the cartoon images and sketches were resized to $192 \times 192$. We split the data into the training and validation sets. As mentioned in Section 6 of the main paper, we also built a test set which consisted of instance sketches from the SketchyScene [4] dataset. The detailed numbers of examples for each category are shown in Table 2.

<table>
<thead>
<tr>
<th>Category</th>
<th>Train Val. Test</th>
<th>Category</th>
<th>Train Val. Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>bench</td>
<td>119 24 50</td>
<td>bird</td>
<td>182 37 100</td>
</tr>
<tr>
<td>bus</td>
<td>167 33 34</td>
<td>butterfly</td>
<td>172 34 50</td>
</tr>
<tr>
<td>car</td>
<td>172 34 150</td>
<td>cat</td>
<td>223 45 50</td>
</tr>
<tr>
<td>chicken</td>
<td>164 33 100</td>
<td>cloud</td>
<td>132 26 50</td>
</tr>
<tr>
<td>cow</td>
<td>178 36 50</td>
<td>dog</td>
<td>165 33 50</td>
</tr>
<tr>
<td>duck</td>
<td>168 34 50</td>
<td>grass</td>
<td>109 22 50</td>
</tr>
<tr>
<td>horse</td>
<td>151 30 50</td>
<td>house</td>
<td>208 41 200</td>
</tr>
<tr>
<td>moon</td>
<td>124 25 50</td>
<td>people</td>
<td>252 51 200</td>
</tr>
<tr>
<td>pig</td>
<td>135 27 50</td>
<td>rabbit</td>
<td>160 32 50</td>
</tr>
<tr>
<td>road</td>
<td>100 20 50</td>
<td>sheep</td>
<td>155 31 50</td>
</tr>
<tr>
<td>star</td>
<td>167 33 50</td>
<td>sun</td>
<td>152 30 50</td>
</tr>
<tr>
<td>tree</td>
<td>139 28 50</td>
<td>truck</td>
<td>128 26 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>3822 765 1734</td>
</tr>
</tbody>
</table>

2. Before collecting the color descriptions, we pre-defined 16 commonly used colors as shown in Table 3, and the semantic part hierarchies for all the 24 categories as in the dataset for instance matching as shown in the “Parts” column in Table 4. We pre-defined the semantic part hierarchies because of the observation that some categories can be entirely described in a single color, while others tend to have different colors for different object parts (e.g., the windows and the body of a car might have different colors). For the latter ones, we need to assign part-level colors.

<table>
<thead>
<tr>
<th>Colors</th>
</tr>
</thead>
<tbody>
<tr>
<td>red, orange, yellow, light green, dark green, cyan, light blue, dark blue, purple, pink, black, light gray, dark gray, light brown, dark brown, white</td>
</tr>
</tbody>
</table>

3. Based on the above preparation for color description collection, we designed an effective approach with the aid of both human manual annotation and automatic generation, which reduced significantly the human effort compared with fully manual annotation.

4. At the human manual annotation side, we designed an easy way for users to make color annotations. For example, to generate the descriptions for the colors of a car and its windows,
we firstly made two folders named with “body” and “windows”. Inside the two folders, we each made 16 empty folders named with the color phrases shown in Table 3. Then, workers only needed to drag-and-drop the collected car images to the 16 empty folders for each part (“body” or “windows”) according to the color of the specified part.

5. At the automatic generation side, we first pre-designed some description patterns for each of the 24 categories according to its semantic part hierarchy, as shown in Table 4. After the human manual annotation, the descriptions were automatically generated with the these sentence patterns.

<table>
<thead>
<tr>
<th>Table 4: Description patterns for foreground categories.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>bench, butterfly,</td>
</tr>
<tr>
<td>cat, cloud, cow,</td>
</tr>
<tr>
<td>dog, duck, grass,</td>
</tr>
<tr>
<td>horse, moon, pig,</td>
</tr>
<tr>
<td>rabbit, road, sheep,</td>
</tr>
<tr>
<td>star, sun, tree</td>
</tr>
<tr>
<td>bird</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>chicken</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>bus</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>car</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>truck</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>house</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>people</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

6. To imitate user inputs in practice, which might contain both location and colorization information, we randomly augmented the location information based on sentence structure patterns for each collected description. For example, in Figure 7 of the main paper, after obtaining “the chicken is light brown” by the above steps, we randomly selected a location phrase from the MATCHING dataset, e.g. “in front of the house”, and inserted it between “the chicken” and “is light brown”. This can be done since we have already known the sentence structures as summarized in Table 4. Thus, we obtained the complete description “the chicken in front of the house is light brown”. Note that this augmentation is optional, because users might not always assign instructions with location information. For example, given a scene sketch with only one car, users probably assign a simple instruction like “the car is/has ...” without describing its location.
With the above procedures, we employed 6 users to annotate, through the drag-and-drop way, the colors of the overall or part-level regions of the cartoon images, and then obtained the description sentences automatically.
2.3 Data Collection for Background Colorization

Figure 1: Illustration of the data collection procedure for background colorization.

The pipeline of the data collection for background colorization is shown in Figure 1 (the same as Figure 8 in the main paper), which produces four modality data: foreground image, background-colorized image, description, and segmentation label map. The detailed procedure is as follows:

1. Since the SketchyScene [4] dataset has provided the ground-truth bounding box (sketch template, Figure 1(a)) of each instance, we first searched our cartoon clip art dataset for the cartoon instances with the same category and similar size to each bounding box and then placed them into a 768 × 768 white canvas, which forms the foreground image, as shown in Figure 1(b).

2. We recruited users to produce the background-colorized images by manually painting the blank regions with solid colors with practical color filling tools such as the Paint tool under Windows. Specifically, we required users to paint with only two colors, “blue” (in RGB (153, 217, 234)) as sky and “green” (in RGB (181, 230, 29)) as ground, as shown in the fourth column of Figure 1.

3. Since we have known the distinct RGB values of the sky and the ground, we obtained the segmentation mask of three categories: sky, ground and foreground simply by checking the color value of each pixel, as shown in Figure 1(c).

4. With the segmentation mask, we first defined several color phrases with different RGB values (11 colors for sky and 5 colors for ground, as shown in Table 5), and then randomly assigned the colors to the sky and ground regions as a data augmentation process for each foreground image. Given the randomly selected colors, we produced the descriptions based on the pattern “the sky is ... and the ground is ...”, as shown in the three columns on the right of Figure 1. Note that the data augmentation and the description generation can both be done automatically, thus making it possible to generate a large-scale dataset.

Table 5: Color definition for background.

<table>
<thead>
<tr>
<th>Sky</th>
<th>red, orange, yellow, green, cyan, blue, purple, pink, black, gray, brown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground</td>
<td>yellow, green, black, gray, brown</td>
</tr>
</tbody>
</table>

With the designed procedures above, we first collected 3932, 300, and 727 sketch templates from the training, validation and test set of the SketchyScene dataset, and then produced foreground images for each template. Afterwards, we employed 24 users to produce a background-colorized image (all in blue sky and green ground) for each foreground image. Finally we automatically augmented each foreground image with 3 more background-colorized images, and totally obtained 15728, 1200, 2908 quadruple data for training, validation, and testing.
3 More Colorization Results

3.1 Un-targeted Colorization

Figure 2 shows more interactive results from the un-targeted colorization study of our system. These results cover instructions with a large degree of diversity, some of which are out of the coverage of the training data mentioned in the main paper, such as “wild” sentence structure (e.g., “light green bus with blue windows” (A3), “red moon in the sky” (E4)), language grammar (e.g., “all the clouds dark gray” (A4), “all the stars is red” (C4)), unsupported words (e.g., the verb “make” in “make the sky blue and ground green” (A5) never appears in the training data).

Figure 2: More interactive results from the un-targeted colorization study of our system.
3.2 Targeted Colorization

Figures 3 to 11 show more results from the targeted colorization study of our system. We invited six users (A: 10-year-old boy in primary school. B: 21-year-old female graduate student. C: 23-year-old male graduate student. D: 22-year-old female graduate student. E: 14-year-old boy in high school. F: 30-year-old female working in a company.) to provide the input instructions for this study. In fact, different users might colorize the targets (foreground objects or background regions) in different orders. While in order to demonstrate the comparisons between instructions towards the same target, we arrange them in the same order. To allow better visualization, we highlight the different expressions towards the same target in red and different color goals in blue.

Figure 3: More results of the targeted colorization study.
References


all the trees are green
all the clouds are green
the grasses are green
the left sheep is light brown
the right sheep is black
the sky is blue and the ground is brown
the leftmost bird is dark blue
the bird on the right most is dark blue
the two middle birds have blue body

Scene sketch
Target cartoon
Output from user D
Output from user F

the trees are all light green
the clouds are all light green
the grasses are all light green
the gray sheep on the left
the black sheep on the left is eating
the sky is blue and the ground is brown
the birds are all blue

Scene sketch
Target cartoon
Output from user C
Output from user E

all the grasses are dark green
all the trees are dark green
the sun is orange
the clouds are light blue
the left butterfly is dark blue
the butterfly on the right is orange
the road is orange
the left dog is brown
the right dog is red
the sky is gray and the ground is yellow

Scene sketch
Target cartoon
Output from user C
Output from user E

grasses are green
all trees are green
sun is yellow
clouds are blue
butterfly on the left is blue
butterfly on the right is orange
road is orange
one dog on the left is brown
the other dog on the right is red
sky is gray and ground is yellow

Figure 4: More results of the targeted colorization study.
the sun in the sky is orange
the clouds are light blue
all the trees are dark green with brown trunks
all the grasses are dark green
the person has a red hair and is in yellow shirt with blue pants
the house is red with gray roof
the dog on the right is dark yellow
the sky is pink and the ground is black

the house is yellow with red roof
one duck on the left is purple
the other duck on the right is white
the road is yellow
trees are green
sun is yellow
grass is dark green
sky is blue and ground is green

Figure 5: More results of the targeted colorization study.
the tree is green
all grass are dark green
the sun is orange
the cloud is blue
the left butterfly is purple
another butterfly on the right is orange
the horse on the left is brown
the pig is gray
the house is yellow with gray roof
the sky is gray and the ground is brown

the road is dark yellow
the stars are yellow
the moon is black
the trees are dark green
the grasses are dark green
the dog is light brown
the house is yellow with red roofs
the sky is blue and the ground is light green

Figure 6: More results of the targeted colorization study.
The sun is orange.

All the trees are dark green.

All the grasses are dark green.

The clouds are gray.

The road is brown.

The left cat is orange.

The cat on the right is cyan.

The sky is pink and the ground gray.

dark yellow dog on right
black person in a suit
blue sky, green ground.

there are two trees, where the leaves are green, the trunks are brown
the sun is orange
green/brown tree

yellow sun
gray cloud
brown sheep
green grass
gray cat

dark yellow dog on right

black person in a suit
blue sky, green ground.

Figure 7: More results of the targeted colorization study.
the bright yellow sun is smiling
the pink clouds are in the sky
the roof of the yellow house is red and the windows are white
the gray dog with dark brown spots is sitting on the road
two yellow chickens are running on the left
two green tree
the duck on the right of the road is red
the grass are green
the sky is cyan and the ground is gray
the clouds are white
all the trees are dark green
the road is brown
the other sheep on the rightmost is black
one sheep on the car is white
two pink cloud
the bus is yellow with blue windows
the grasses are green
the sky is black and the ground is light green

Figure 8: More results of the targeted colorization study.
Scene sketch

<table>
<thead>
<tr>
<th>Input from user B</th>
<th>Output from user D</th>
</tr>
</thead>
<tbody>
<tr>
<td>the sun is orange</td>
<td>the sun is orange</td>
</tr>
<tr>
<td>the light blue clouds</td>
<td>the cloud is light blue</td>
</tr>
<tr>
<td>the rightmost bird under sun is blue</td>
<td>the right bird is light blue</td>
</tr>
<tr>
<td>another bird at left is yellow</td>
<td>the bird in the middle is yellow</td>
</tr>
<tr>
<td>the butterfly is orange</td>
<td>the butterfly is orange</td>
</tr>
<tr>
<td>draw the tree light green</td>
<td>all the trees are green</td>
</tr>
<tr>
<td>the road is black</td>
<td>the road is black</td>
</tr>
<tr>
<td>the car on the left is yellow with the black windows</td>
<td>the left car is yellow with black windows</td>
</tr>
<tr>
<td>the other car on the right is blue and white, with the light blue windows</td>
<td>the right car is dark blue with light blue windows</td>
</tr>
<tr>
<td>the grasses are green</td>
<td>all the grasses are green</td>
</tr>
<tr>
<td>draw the sky blue and ground light green</td>
<td>blue sky and green ground</td>
</tr>
</tbody>
</table>

Scene sketch

<table>
<thead>
<tr>
<th>Input from user B</th>
<th>Output from user A</th>
<th>Output from user D</th>
</tr>
</thead>
<tbody>
<tr>
<td>yellow car</td>
<td>the car is yellow with dark blue windows</td>
<td>all the trees are dark green</td>
</tr>
<tr>
<td>green/brown tree</td>
<td>all the trees are dark green</td>
<td>the cloud is gray</td>
</tr>
<tr>
<td>gray cloud</td>
<td>the cloud is gray</td>
<td>all the grasses are light green</td>
</tr>
<tr>
<td>green grass</td>
<td>the sun is yellow</td>
<td>yellow sun</td>
</tr>
<tr>
<td>yellow sun</td>
<td>the sun is yellow</td>
<td>purple butterfly</td>
</tr>
<tr>
<td>purple butterfly</td>
<td>the butterfly is purple</td>
<td>black road</td>
</tr>
<tr>
<td>black road</td>
<td>the road is black</td>
<td>sky is orange. ground is yellow</td>
</tr>
<tr>
<td>sky is orange. ground is yellow.</td>
<td>the sky is orange and the ground is yellow</td>
<td></td>
</tr>
</tbody>
</table>
the sun is orange
the clouds are black
draw the house pink, and it has
the purple roof
the butterfly is blue
the cow in the middle is blue
the truck is red in the front, and
the behind is white
the tree is green
also the grass is green
the other cow in the front is
dark brown
the sheep is pink
draw the sky blue and the ground is gray

the sun near the cloud is orange
the cloud is black
draw the house pink, and it has
the purple roof
the butterfly is blue
the cow in the middle is blue
the truck is red in the front, and
the behind is white
the tree is green
also the grass is green
the other cow in the front is
dark brown
the sheep is pink
draw the sky blue and the ground is gray

the sun is orange
the clouds are black
the house is pink with purple
roof
the flying butterfly is blue
the cow behind the truck is blue
the truck has a red headstock
with cyan window and has a
gray body
the tree is dark green with
brown trunk
all the grasses are green
one cow in the lower left corner
is light brown
the sheep is light brown
the sky is light blue and the
ground is gray

Figure 10: More results of the targeted colorization study.
some light blue clouds are floating in the sky
four green trees stand on the ground
the grasses are dark green
the orange sun has a glass
the dog near the house is light brown
the dog on the leftmost is light brown with some dark grown spots
the dog on the bottom is totally dark brown
the light blue bird on the left has a pair of dark blue wings
the light green bird flying on the right has a pair of dark green wings
the house is yellow, and the roof is red
the sky is purple and the ground is gray.

Figure 11: More results of the targeted colorization study.