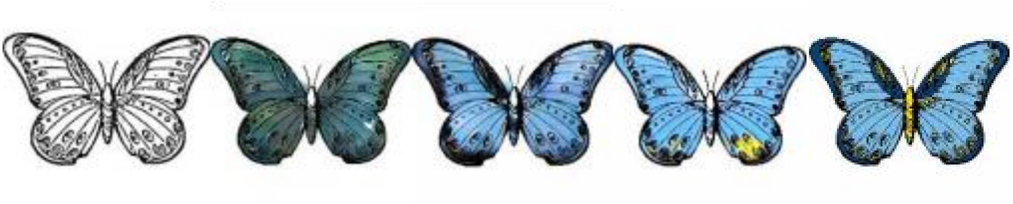
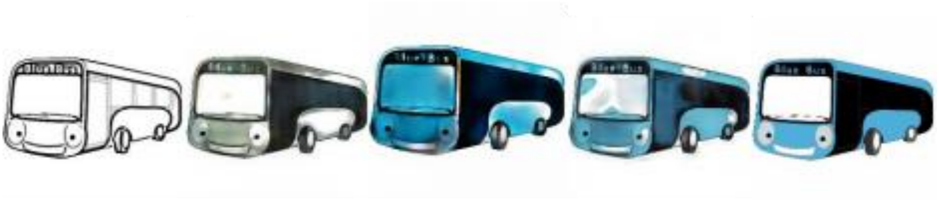
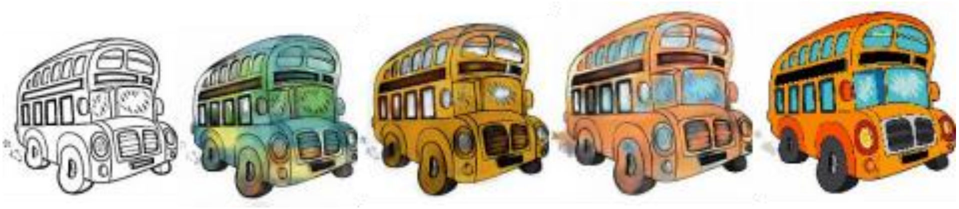
The butterfly is blue with yellow spots.



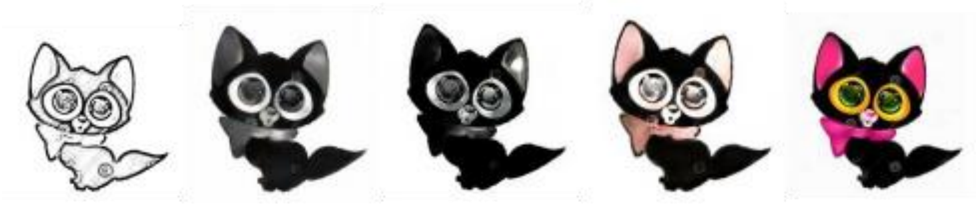
The bus is blue with white window.



The bus is yellow with blue window.



The cat is black with pink ear and Scarf.



The forground colouring results are shown in Figure. A total of three models were trained for comparison during the experimental phase. (a) is the structure proposed by Zou et al.\cite{zou2019language} using LSTM for text encoding and for text image fusion. (b) is the original version was replaced using GRU and adjusting the learning rate to 0.00002, and (c) the structure proposed in Chapter 3 using Bert for text encoding and residual block for text image fusion.

As shown in the Figure, the coloring of the wings of the butterfly with GRU for text encoding and text image fusion and with Bert for text encoding and fusion of text image information using the residual block is closer to the ground truth than the structure with LSTM for text encoding and fusion of text image information. and compared to the structure with LSTM for text encoding and fusion of text image information, the coloring of the wings of the butterfly with GRU for text encoding and fusion of text image information is closer to the ground truth. The model proposed in Chapter 3, which uses Bert for text encoding and a residual block for text image fusion, shows the detail of the butterfly's yellow spots, which is better than the model proposed in Chapter 3 in terms of detailing and taking into account multiple colours when multiple colours are present.

[1] Zou, C., Mo, H., Gao, C., Du, R., & Fu, H. (2019). Language-based colorization of scene sketches. ACM Transactions on Graphics (TOG), 38(6), 1-16.