Abstract
We investigate different methods to generate colorful cartoon images from black and white sketches, together with the color hints given by user. An end-to-end method is firstly implemented using CNN with direct links, namely uNet. We also try conditional generative adversarial networks (cGAN) [1] [2], Wasserstein GAN (WGAN) [3] and improved WGAN [4] to improve the generating quality.

Method
Wasserstein GAN value function:
\[
\min_{G} \max_{D} \mathbb{E}_{x \sim p_{data}} [D(x)] - \mathbb{E}_{\tilde{x} \sim p_{G}} [D(\tilde{x})]
\]
Improved-WGAN training strategy: We follow [4] to use gradient penalty to enforce the Lipschitz constraint. The objective function is as follows:
\[
L = \mathbb{E}_{\tilde{x} \sim p_{G}} [D(\tilde{x})] - \mathbb{E}_{x \sim p_{data}} [D(x)] + \lambda \mathbb{E}_{\tilde{x} \sim p_{G}} [\|\nabla_{\tilde{x}} D(\tilde{x})\|_{2} - 1]^{2}.
\]
L1 loss: Previous approaches of c-GANs [2] have found it beneficial to mix the GAN objective with a more conventional loss functions. In this paper, we also use the L1 distance to describe the pixel-level loss in our model.

Network Structure
Generator: The generator produces a colored image based on line image and color hints. Instead of the encoder-decorder structure, we employ the "U-Net" [5] by concatenating layers in encoder to the corresponding layers of the decoder. The network structure is in Fig 2.

Problem Statement
Given color hint and line art image, we colorize the sketch.
Dataset: 20000 colored manga images from safebooru.org

Quantitative Results
Fig 4. Improved-WGAN discriminator loss curve.
(-1 * d_loss) represents the W-distance. Smaller the distance represents higher similarity between the generated images and real images. From Fig 4 we can see that the discriminator's loss steadily go up until convergence.

Experiment Results
Fig 6. Experiment Results on Test Set

References