Term Project Final Presentation

Computer Vision

| 2022.06.07 | Team 1

14446_001, Tue/Wed
Department of software

201533631 김도균
201735900 전탁
201935017 김민경
201935048 박준이
201935034 김한결
201935135 조수현

Ⅰ. Introduction

Ⅰ. What is ConvNet Challenge

Purpose of the Challenge

To restore the color through model training the gray tone image and the hint color map for that image are given.

Ⅰ. What is ConvNet Challenge

Purpose of the Challenge

When an image with a Gray Scale shape and a corresponding color-hint map are given, the task of restoring colors is performed.

Ⅱ. Related Work (Paper)

Ⅱ. Real-Time User-Guided Image Colorization with Learned Deep Priors

Real-Time User-Guided Image Colorization with Learned Deep Priors

Richard Zhang* Jun-Yan Zhu* Phillip Isola Xinyang Geng Angela S. Lin Tianhe Yu Alexei A. Efros
University of California, Berkeley

Code [GitHub] [Training] SIGGRAPH 2017 [Paper] Conference [Talk] Slides [ppt]

Abstract

We propose a deep learning approach for user-guided image colorization. The system directly maps a grayscale image, along with sparse, local user ``hints" to an output colorization with a Convolutional Neural Network (CNN). Rather than using hand-defined rules, the network propagates user edits by fusing low-level cues along with high-level semantic information, *learned from large-scale data*. We train on a million images, with simulated user inputs. To guide the user towards efficient input selection, the system recommends likely colors based on the input image and current user inputs. The colorization is performed in a single feed-forward pass, enabling real-time use. Even with randomly simulated user inputs, we show that the proposed system helps novice users quickly create realistic colorizations, and show large improvements in colorization quality with just a minute of use. In addition, we show that the framework can incorporate other user "hints" as to the desired colorization, showing an application to color histogram transfer.

* https://richzhang.github.io/InteractiveColorization/

Propose a deep learning approach for user-guided image colorization.

To guide the user towards efficient input selection, the system recommends likely colors based on the input image and current user inputs.

Show that the framework can incorporate other user "hints" as to the desired colorization, showing an application to color histogram transfer.

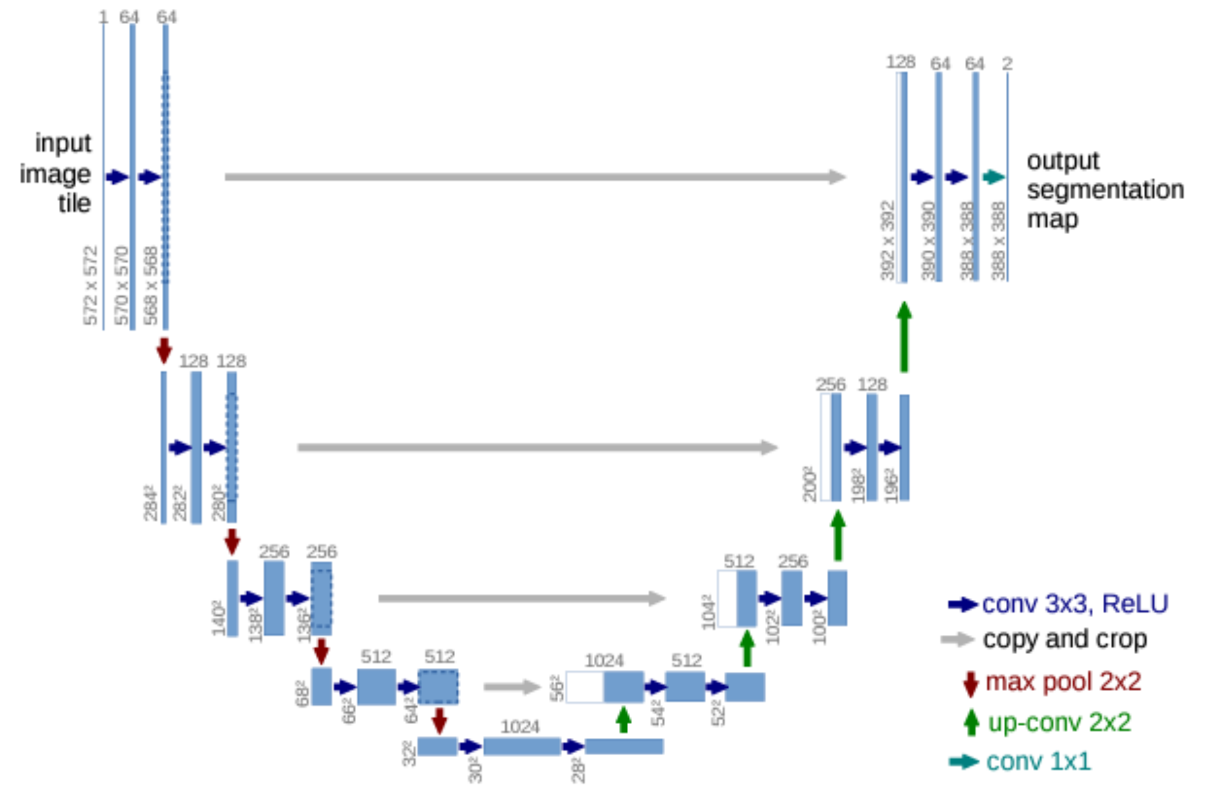
Ⅱ. U-Net: Convolutional Networks for Biomedical Image Segmentation

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, and Thomas Brox

Computer Science Department and BIOSS Centre for Biological Signalling Studies,
University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/

Abstract. There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. Using the same network trained on transmitted light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover, the network is fast. Segmentation of a 512x512 image takes less than a second on a recent GPU. The full implementation (based on Caffe) and the trained networks are available at http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net.



* https://arxiv.org/abs/1505.04597

Ⅱ. Road Extraction by Deep Residual U-Net

Road Extraction by Deep Residual U-Net

Zhengxin Zhang†, Qingjie Liu†*, *Member, IEEE* and Yunhong Wang, *Senior Member, IEEE*

***Abstract*—Road extraction from aerial images has been a hot research topic in the field of remote sensing image analysis. In this letter, a semantic segmentation neural network which combines the strengths of residual learning and U-Net is proposed for road area extraction. The network is built with residual units and has similar architecture to that of U-Net. The benefits of this model is two-fold: first, residual units ease training of deep networks. Second, the rich skip connections within the network could facilitate information propagation, allowing us to design networks with fewer parameters however better performance. We test our network on a public road dataset and compare it with U-Net and other two state of the art deep learning based road extraction methods. The proposed approach outperforms all the comparing methods, which demonstrates its superiority over recently developed state of the arts.**

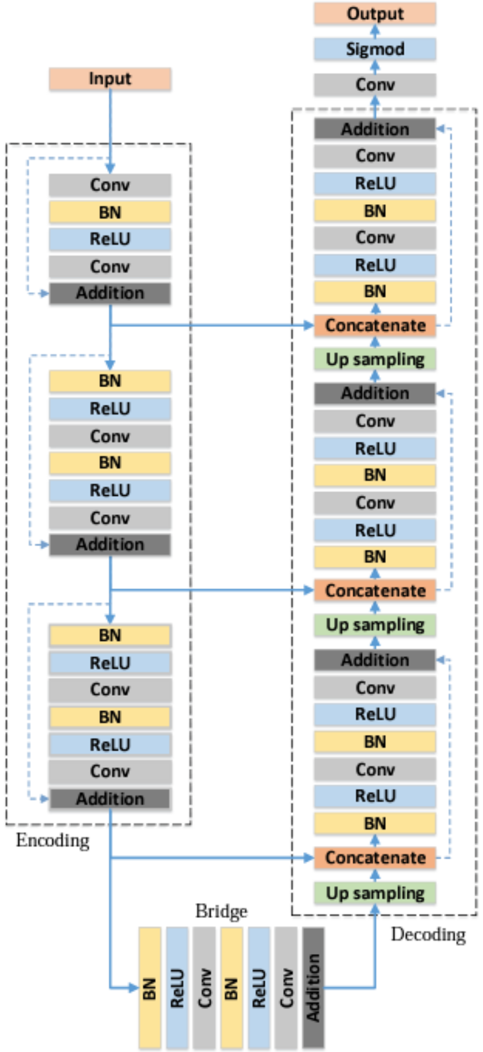
***Index Terms*—Road extraction, Convolutional Neural Network, Deep Residual U-Net.**

multi-spectral images using probabilistic SVM. Alshehhi and Marpu [6] proposed an unsupervised road extraction method based on hierarchical graph-based image segmentation.

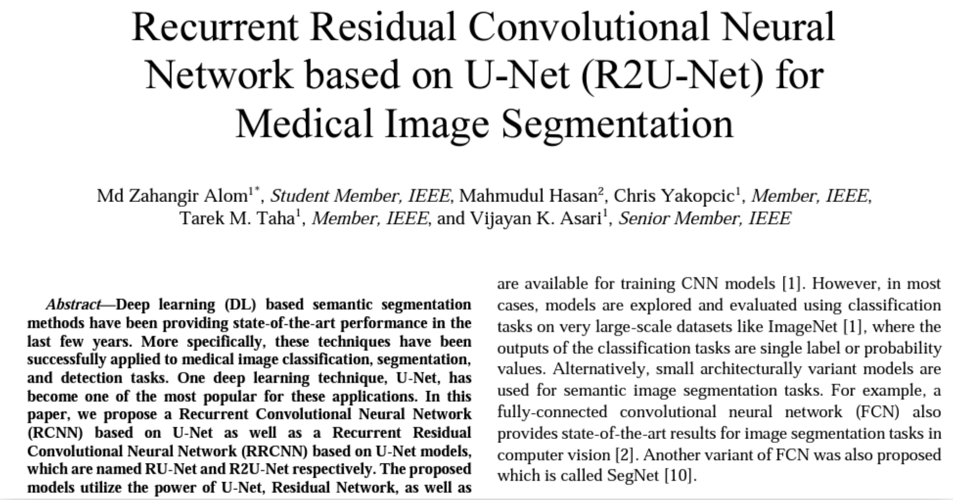
Recent years have witnessed great progress in deep learning. Methods based on deep neural networks have achieved state-of-the-art performance on a variety of computer vision tasks, such as scene recognition [13] and object detection [14]. Researchers in remote sensing community also seek to leverage the power of deep neural networks to solve the problems of interpretation and understanding of remote sensing data [2], [5], [15]–[18]. These methods provide better results than traditional ones, showing great potential of applying deep learning techniques to analyze remote sensing tasks.

In the field of road extraction, one of the first attempts of applying deep learning techniques was made by Mnih and

* https://arxiv.org/abs/1711.10684



Ⅱ. Recurrent Residual Convolutional Neural Network based on U-Net

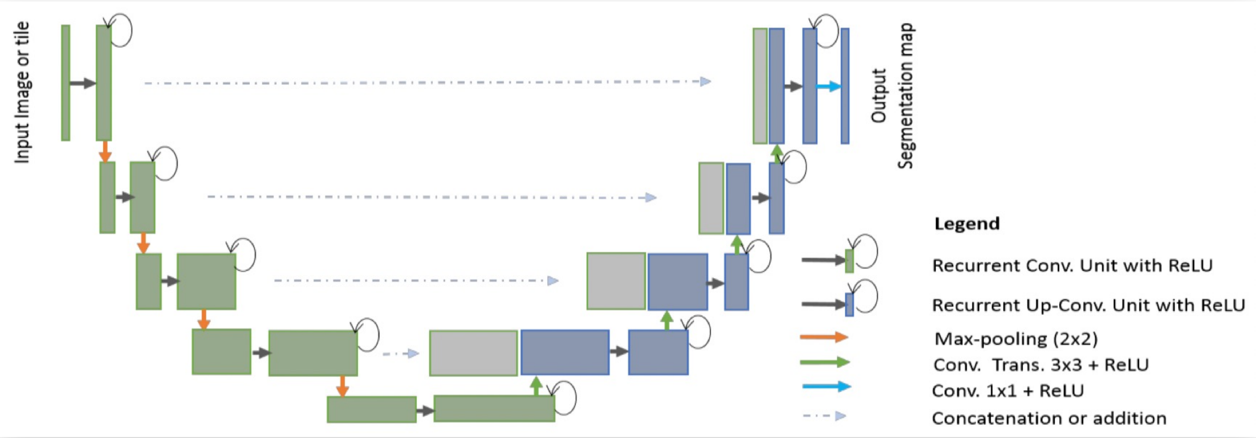


Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation

Md Zahangir Alom[1*], *Student Member, IEEE,* Mahmudul Hasan[2], Chris Yakopcic[1], *Member, IEEE,* Tarek M. Taha[1], *Member, IEEE,* and Vijayan K. Asari[1], *Senior Member, IEEE*

***Abstract*—Deep learning (DL) based semantic segmentation methods have been providing state-of-the-art performance in the last few years. More specifically, these techniques have been successfully applied to medical image classification, segmentation, and detection tasks. One deep learning technique, U-Net, has become one of the most popular for these applications. In this paper, we propose a Recurrent Convolutional Neural Network (RCNN) based on U-Net as well as a Recurrent Residual Convolutional Neural Network (RRCNN) based on U-Net models, which are named RU-Net and R2U-Net respectively. The proposed models utilize the power of U-Net, Residual Network, as well as**

are available for training CNN models [1]. However, in most cases, models are explored and evaluated using classification tasks on very large-scale datasets like ImageNet [1], where the outputs of the classification tasks are single label or probability values. Alternatively, small architecturally variant models are used for semantic image segmentation tasks. For example, a fully-connected convolutional neural network (FCN) also provides state-of-the-art results for image segmentation tasks in computer vision [2]. Another variant of FCN was also proposed which is called SegNet [10].



* https://arxiv.org/abs/1802.06955

Ⅱ. Attention U-Net: Learning Where to Look for the Pancreas

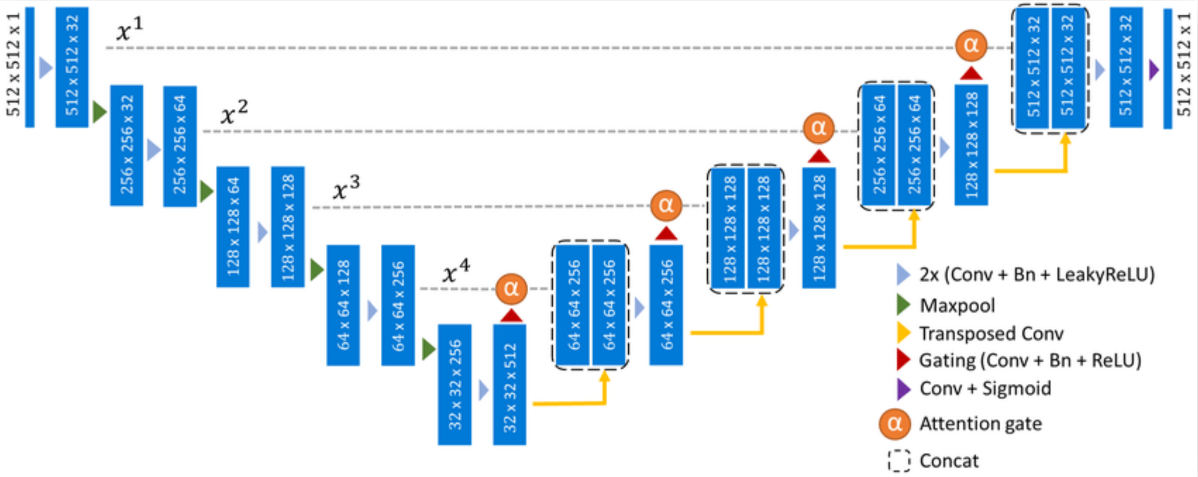
Attention U-Net: Learning Where to Look for the Pancreas

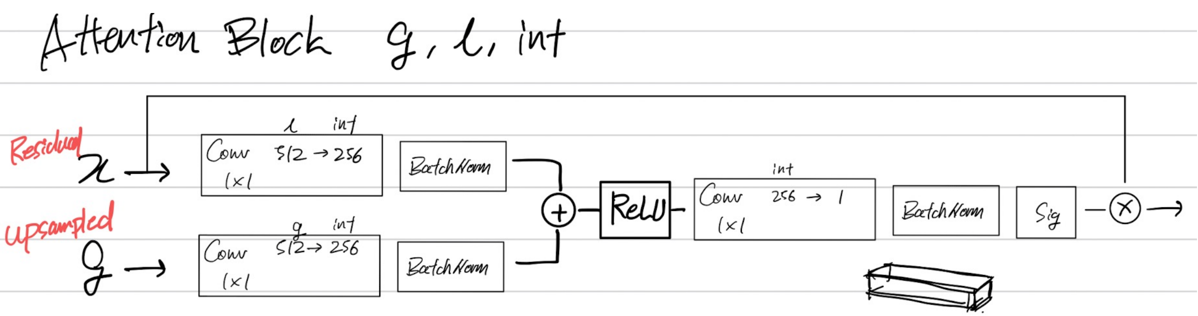
Ozan Oktay[1,5], Jo Schlemper[1], Loic Le Folgoc[1], Matthew Lee[4], Mattias Heinrich[3], Kazunari Misawa[2], Kensaku Mori[2], Steven McDonagh[1], Nils Y Hammerla[5], Bernhard Kainz[1], Ben Glocker[1], and Daniel Rueckert[1]

[1]Biomedical Image Analysis Group, Imperial College London, London, UK
[2]Dept. of Media Science, Nagoya University & Aichi Cancer Center, JP
[3]Medical Informatics, University of Luebeck, DE, [4]HeartFlow, California, USA
[5]Babylon Health, London, UK

Abstract

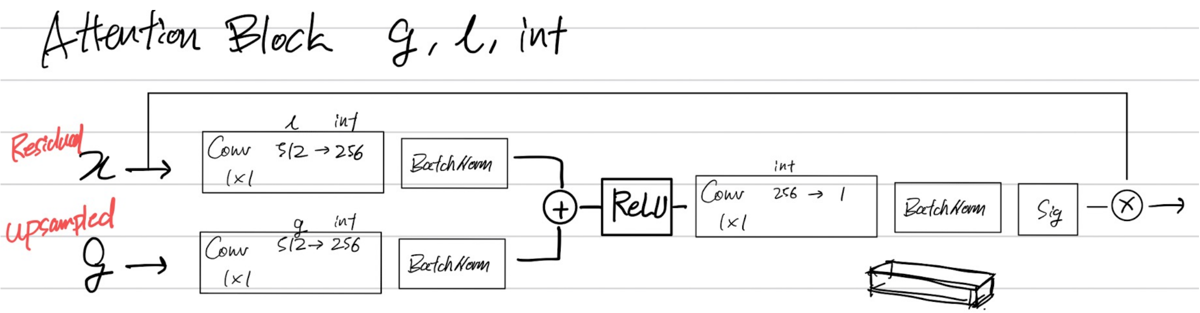
We propose a novel attention gate (AG) model for medical imaging that automatically learns to focus on target structures of varying shapes and sizes. Models trained with AGs implicitly learn to suppress irrelevant regions in an input image while highlighting salient features useful for a specific task. This enables us to eliminate the necessity of using explicit external tissue/organ localisation modules of cascaded convolutional neural networks (CNNs). AGs can be easily integrated into standard CNN architectures such as the U-Net model with minimal computational overhead while increasing the model sensitivity and prediction accuracy. The proposed Attention U-Net architecture is evaluated on two large CT abdominal datasets for multi-class image segmentation. Experimental results show that AGs consistently improve the prediction performance of U-Net across different datasets and training sizes while preserving computational efficiency. The source code for the proposed architecture is publicly available.

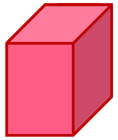


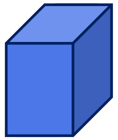


* https://arxiv.org/abs/1804.03999

Ⅱ. Attention U-Net: Learning Where to Look for the Pancreas

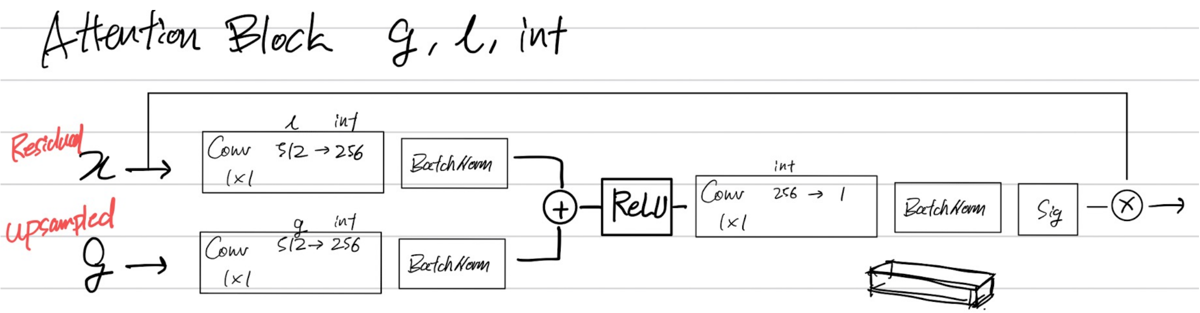


Residual 

Up sampled 

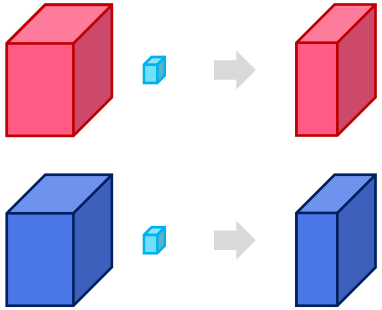
* https://arxiv.org/abs/1804.03999

Ⅱ. Attention U-Net: Learning Where to Look for the Pancreas



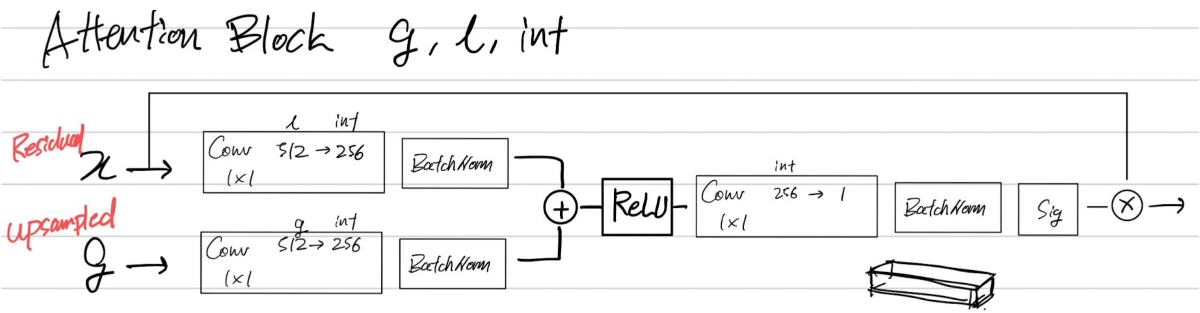
Residual

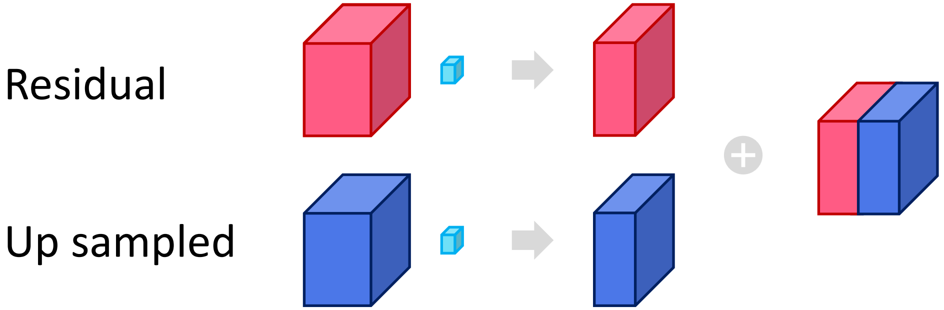
Up sampled



* https://arxiv.org/abs/1804.03999

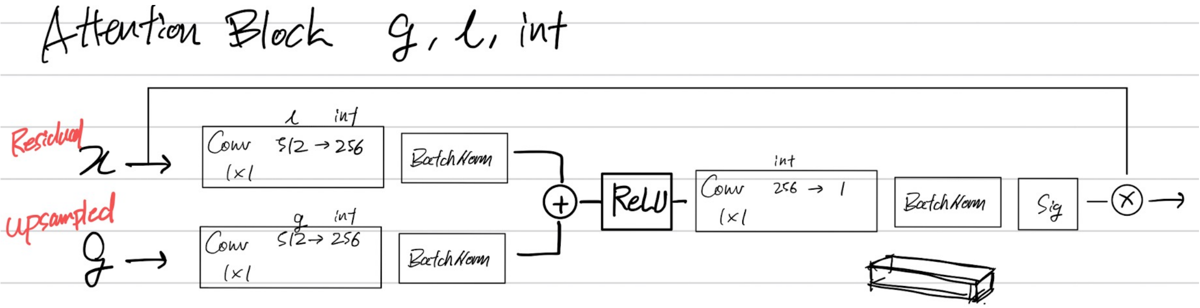
Ⅱ. Attention U-Net: Learning Where to Look for the Pancreas





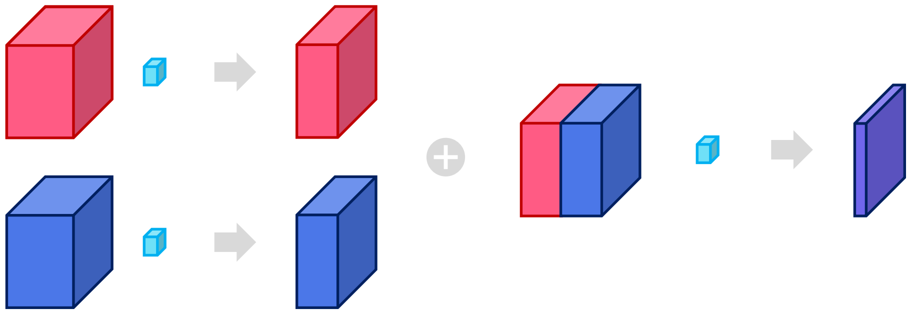
* https://arxiv.org/abs/1804.03999

Ⅱ. Attention U-Net: Learning Where to Look for the Pancreas



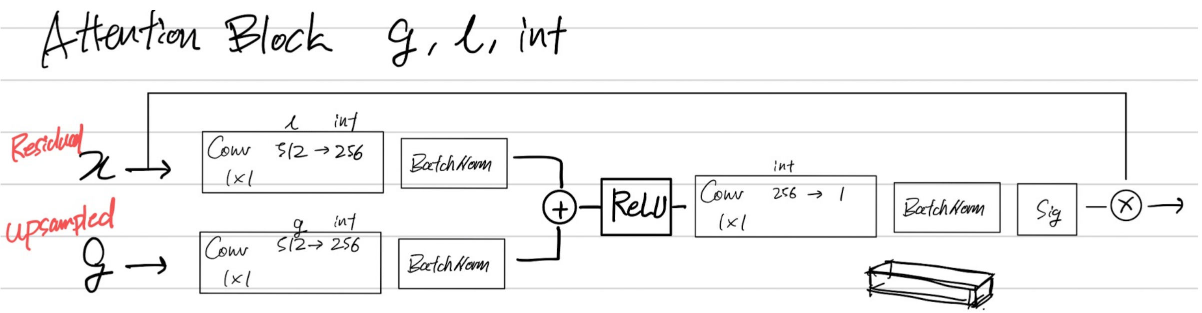
Residual

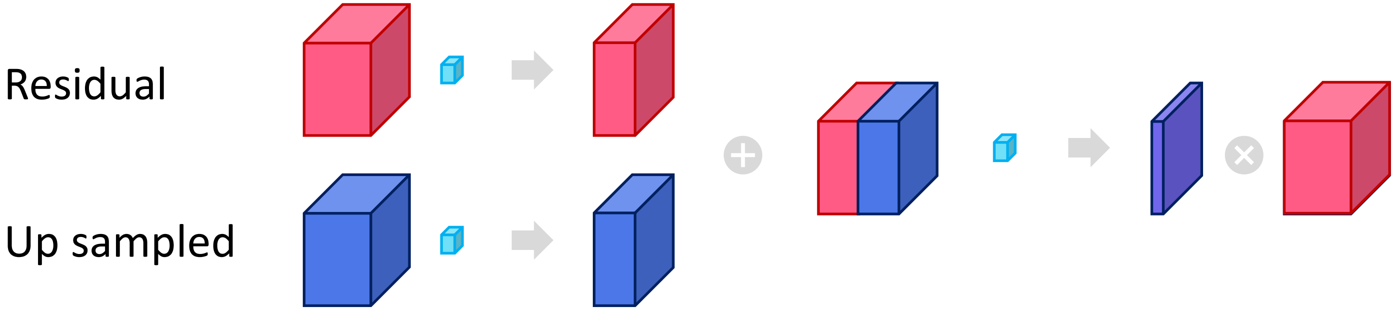
Up sampled



* https://arxiv.org/abs/1804.03999

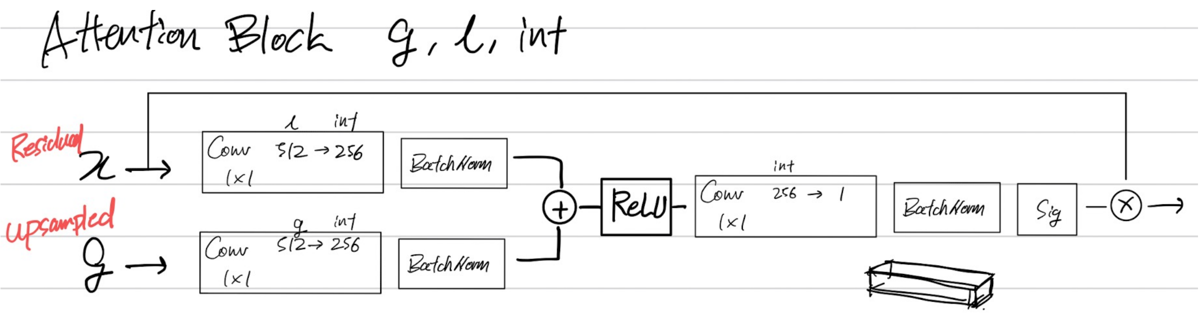
Ⅱ. Attention U-Net: Learning Where to Look for the Pancreas





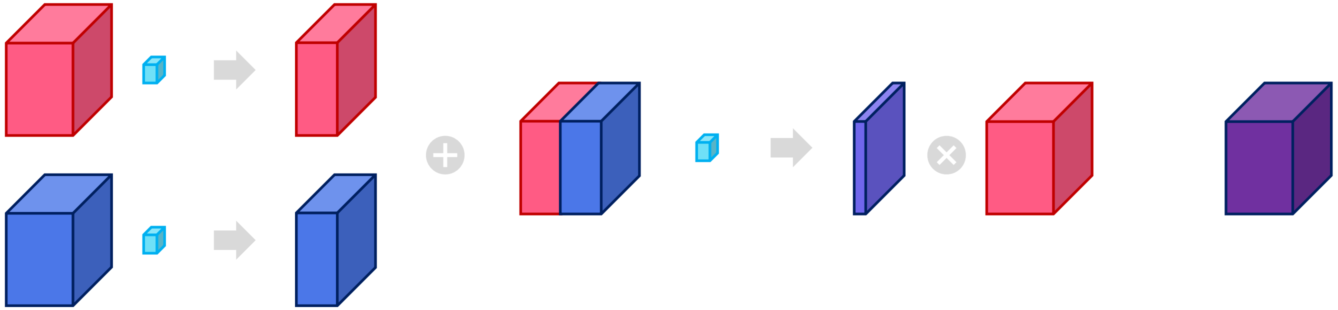
* https://arxiv.org/abs/1804.03999

Ⅱ. Attention U-Net: Learning Where to Look for the Pancreas



Residual

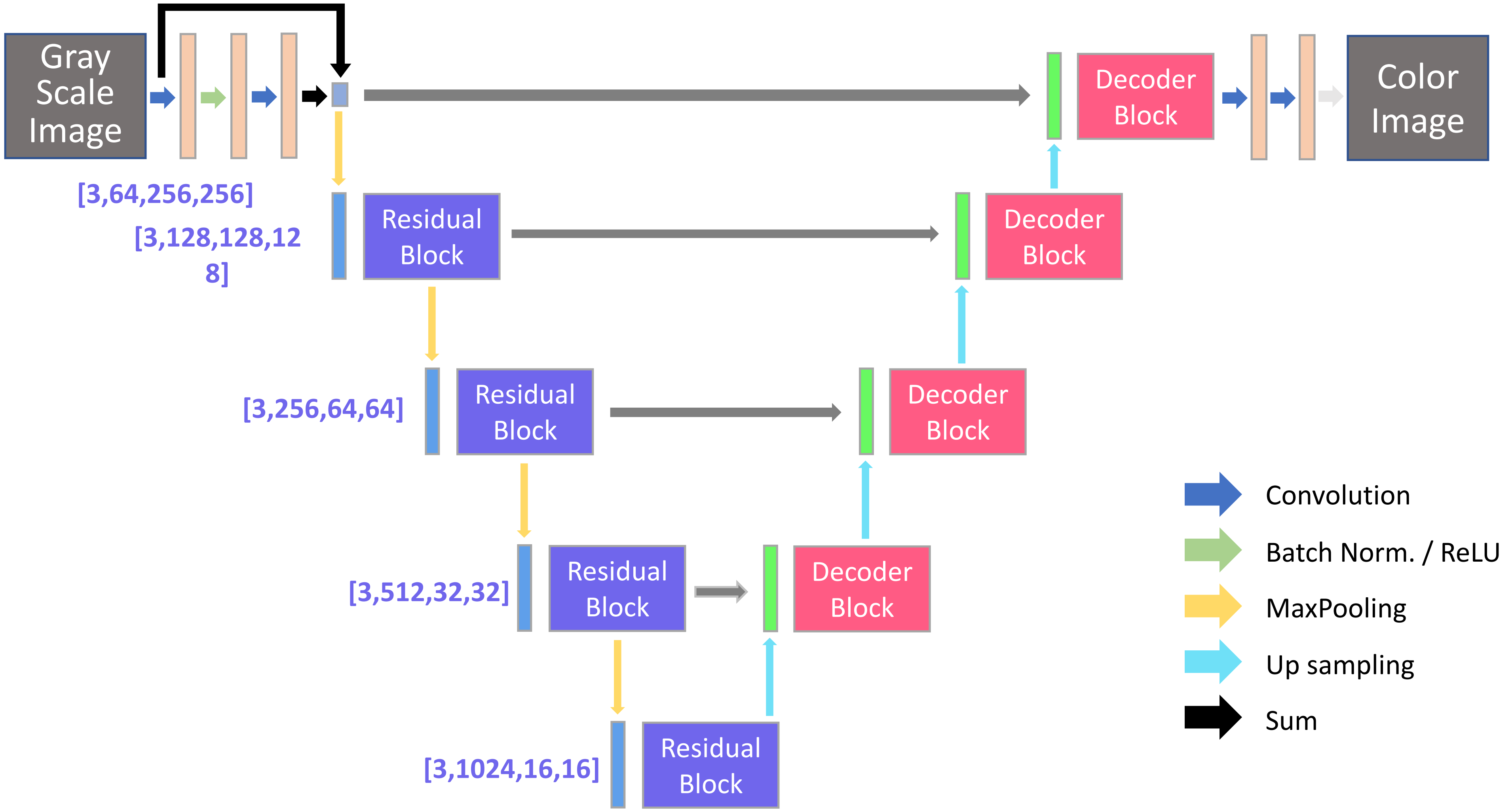
Up sampled



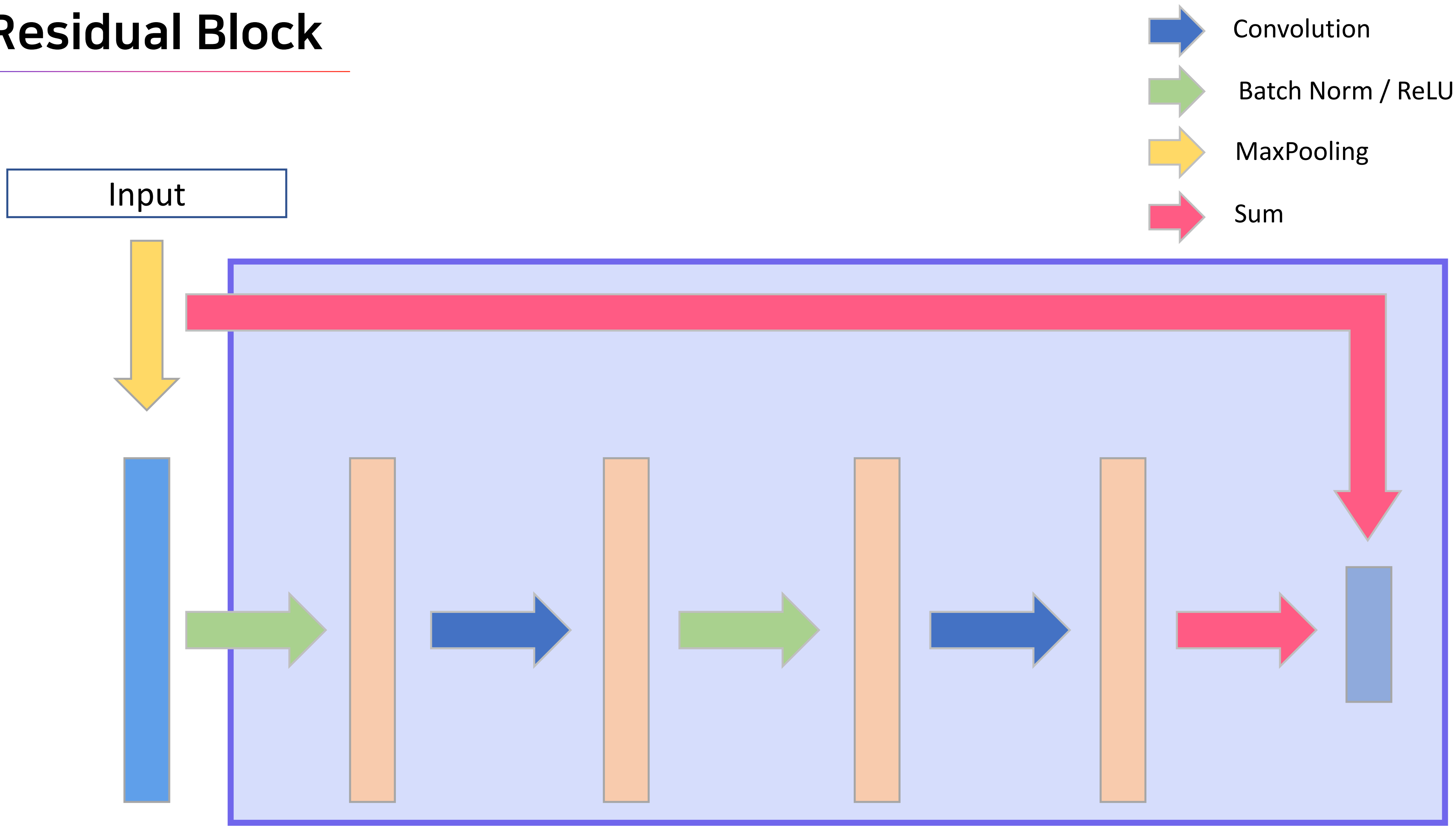
* https://arxiv.org/abs/1804.03999

Ⅲ. Model Overview

Ⅲ. Model Overview

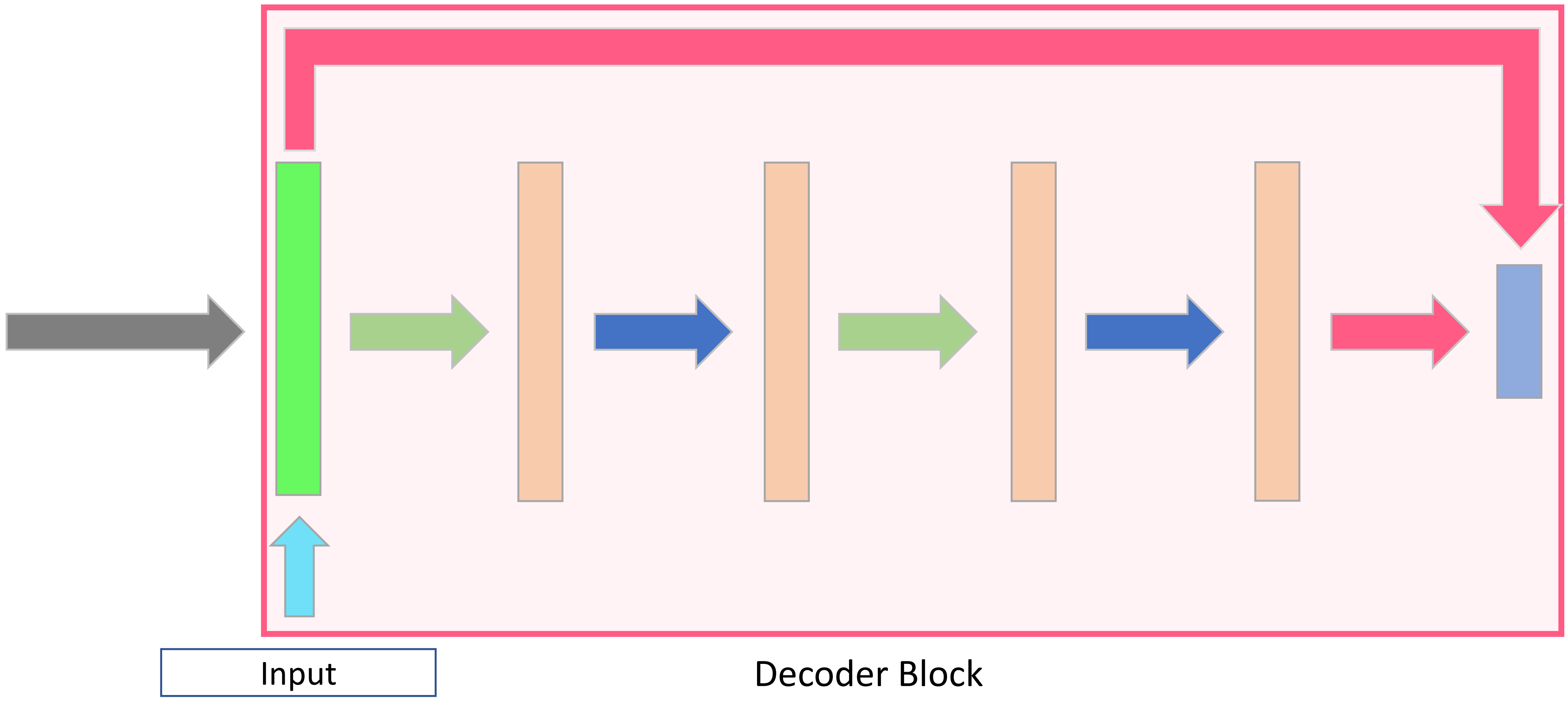


Ⅲ. Residual Block

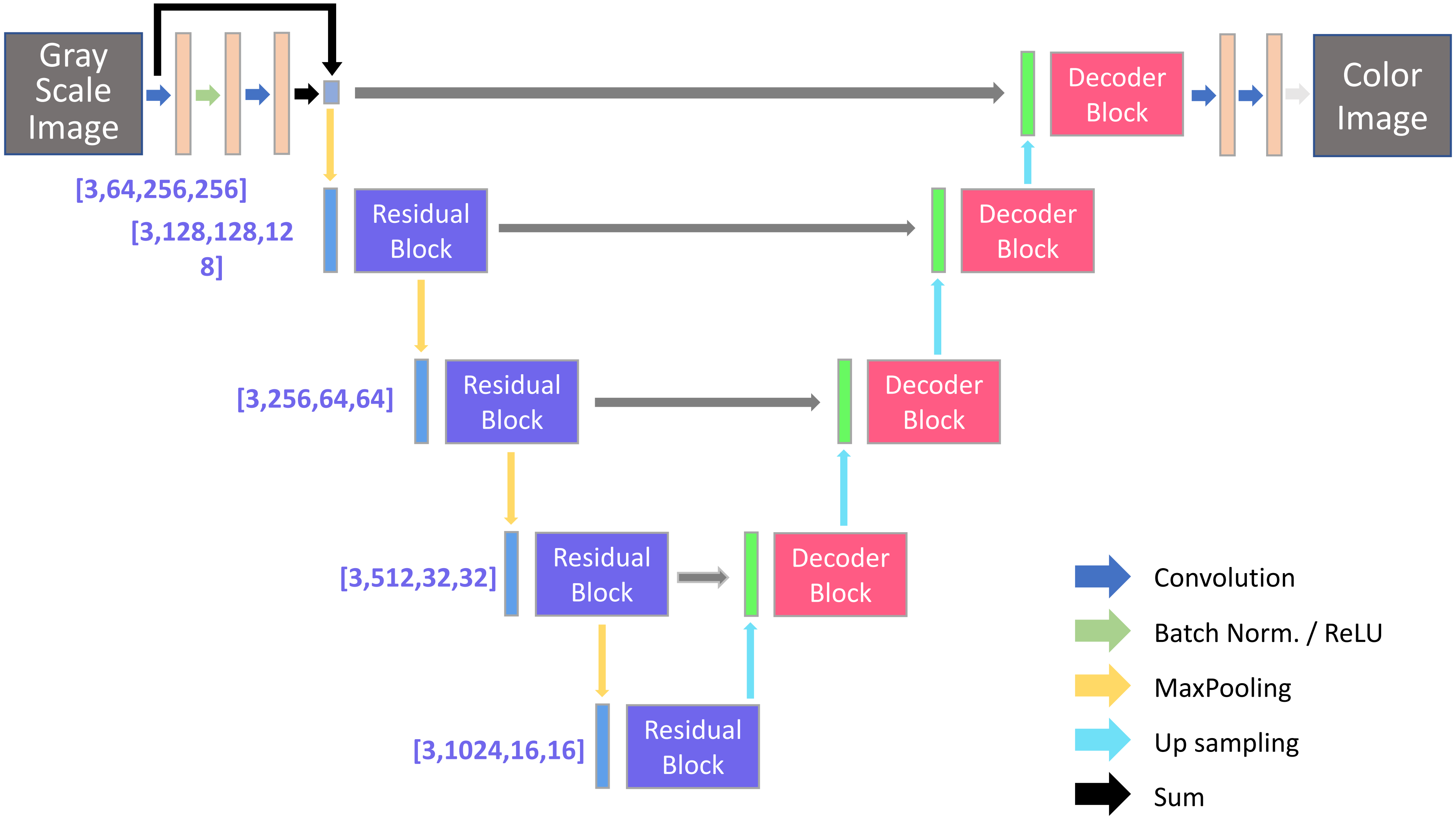


Residual Block

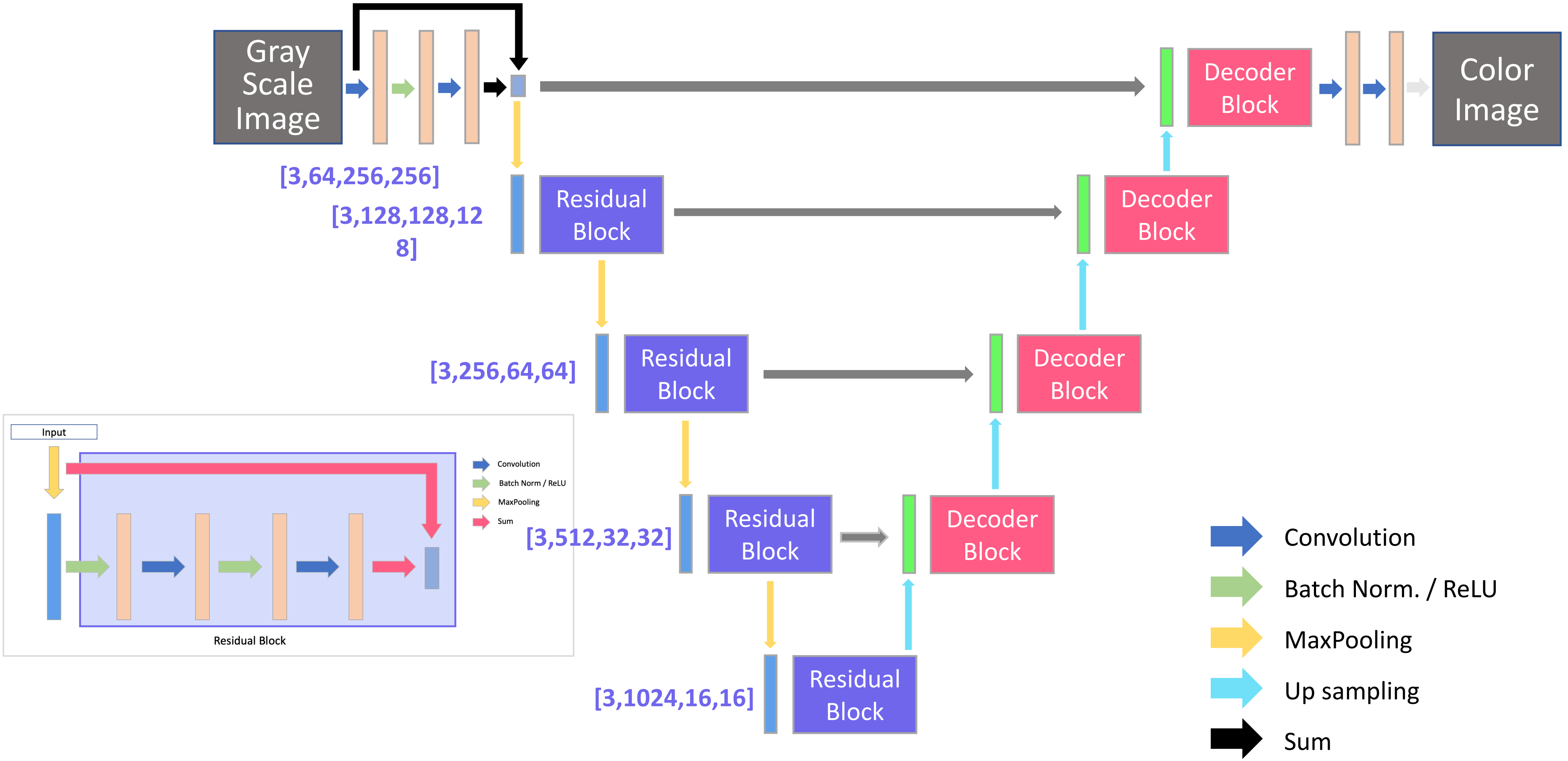
Ⅲ. Decoder Block



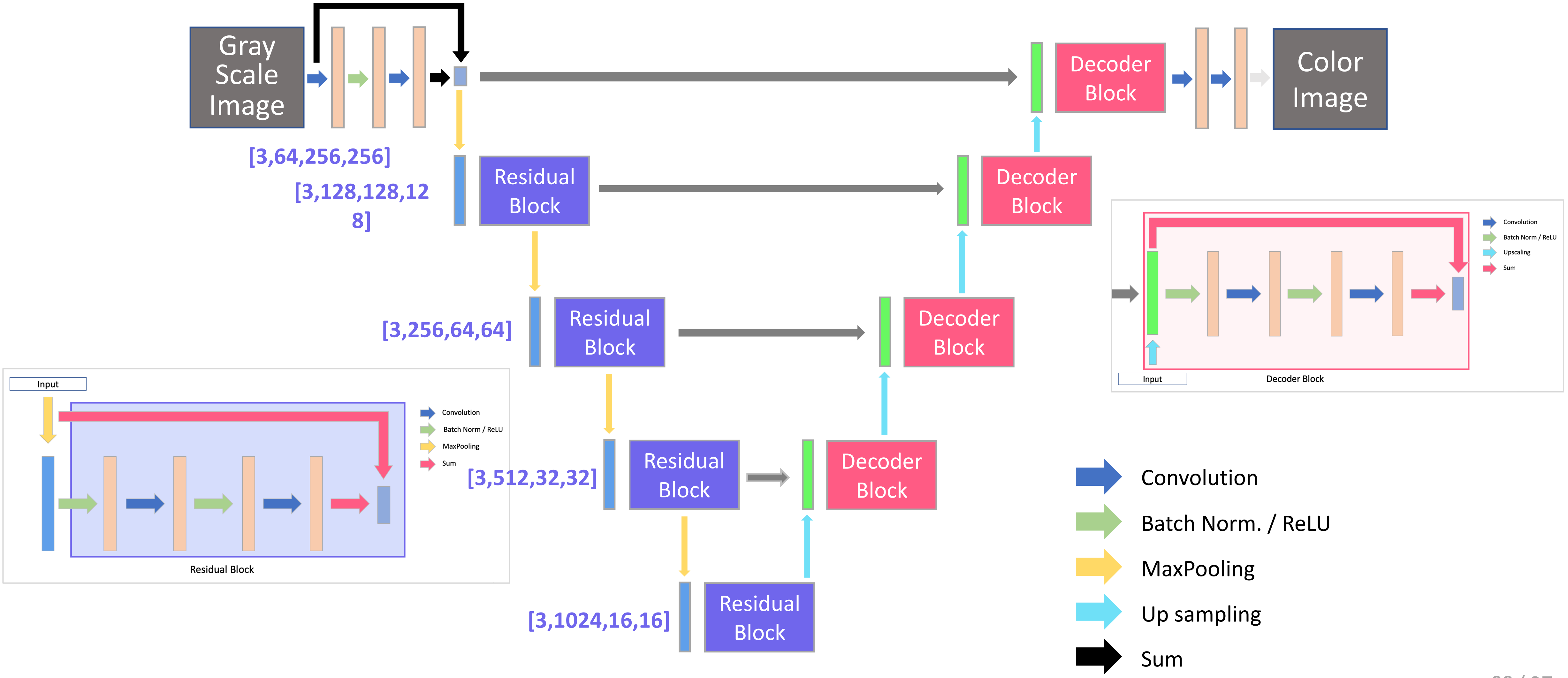
Ⅲ. Model Overview



Ⅲ. Model Overview

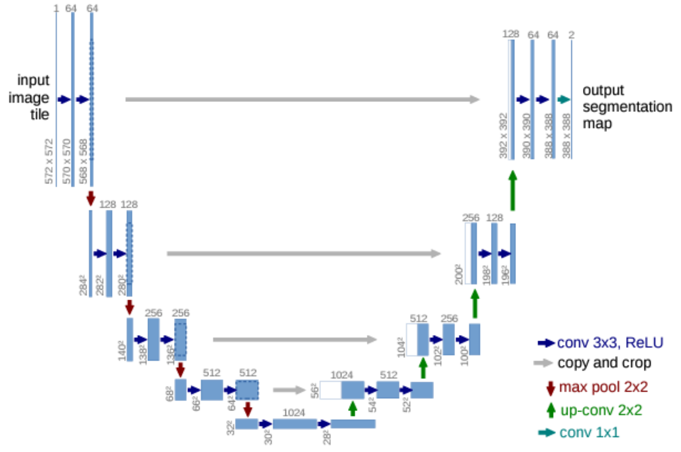


Ⅲ. Model Overview

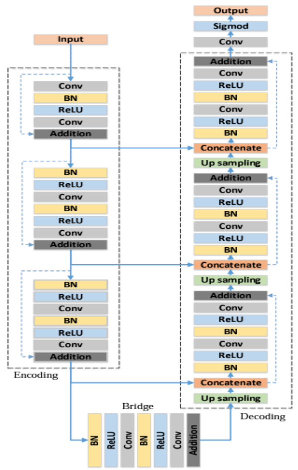


Ⅳ. Experiments

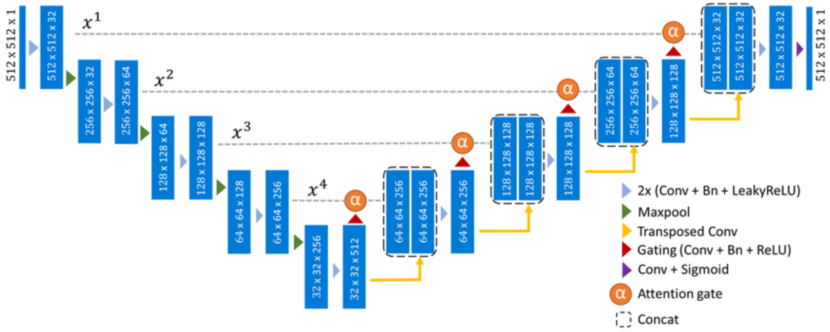
Ⅳ. Tested Models



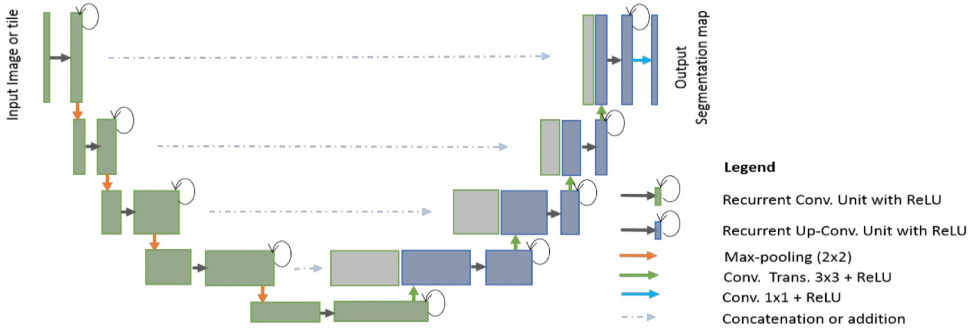
U-net



Residual U-Net



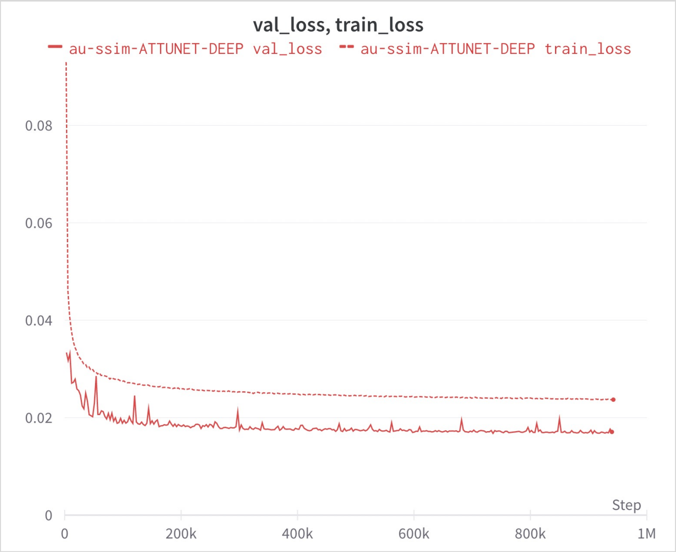
Attention U-Net



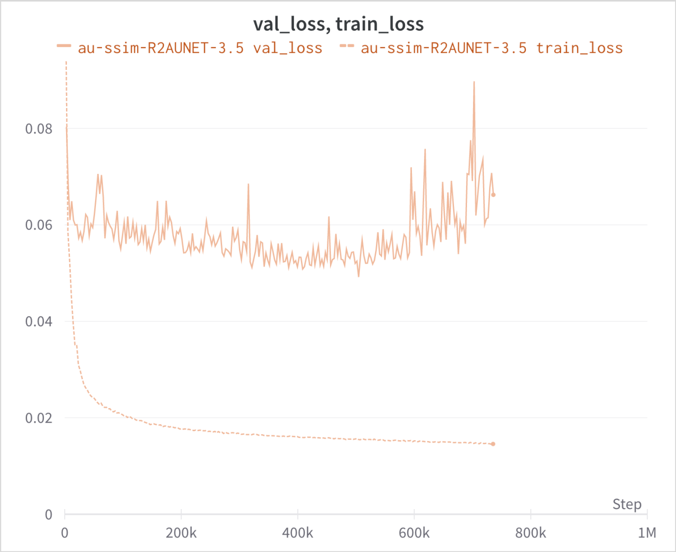
Recurrent Residual U-Net

Ⅳ. Overfitted in R2AU-Net

Residual Attention U-Net



Residual Recurrent Attention U-Net



* Training epoch: 245

Ⅳ. Loss Function

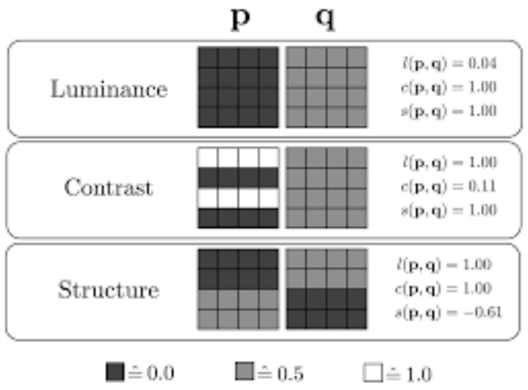
Multi Scale SSIM Loss Function

$$SSIM(x, y) = [l(x, y)]^{\alpha} \cdot [c(x, y)]^{\beta} \cdot [s(x, y)]^{\gamma}$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$



`C1 = (0.01*L).^2`, where `L` is the specified `DynamicRange` value.

`C2 = (0.03*L).^2`, where `L` is the specified `DynamicRange` value.

`C3 = C2/2`

$$\begin{aligned} \mathrm{SSIM}(p) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ &= l(p) \cdot cs(p) \end{aligned}$$

$$\text{MS-SSIM}(p) = l_M^{\alpha}(p) \cdot \prod_{j=1}^{M} cs_j^{\beta_j}(p)$$

$$= [l_M(\mathbf{x}, \mathbf{y})]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(\mathbf{x}, \mathbf{y})]^{\beta_j} [s_j(\mathbf{x}, \mathbf{y})]^{\gamma_j}$$

Code Reference

https://github.com/psyrocloud/MS-SSIM_L1_LOSS

Ⅳ. Loss Function

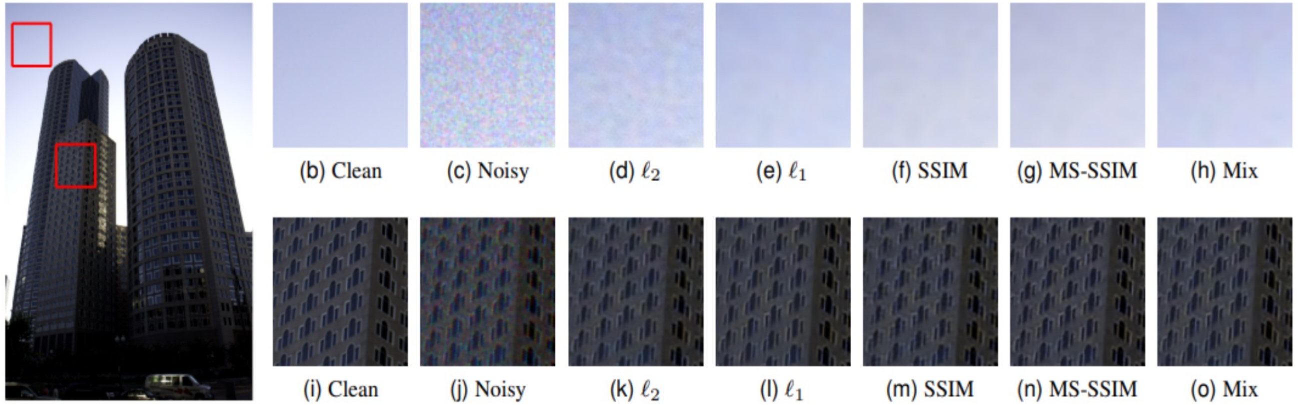
L1+MS SSIM Loss Function

- Formula = **α * MS_SSIM + (1- α) * L1**
- Alpha α = **0.025**
- We put more weight on L1, and as a result, we derived good performance.

Denoising + demosaicking			Training cost function					
Image quality metric	Noisy	*BM3D*	ℓ_2	ℓ_1	$SSIM_5$	$SSIM_9$	MS-SSIM	Mix
$1000 \cdot \ell_2$	1.65	0.45	0.56	0.43	0.58	0.61	0.55	**0.41**
PSNR	28.24	34.05	33.18	34.42	33.15	32.98	33.29	**34.61**
$1000 \cdot \ell_1$	27.36	14.14	15.90	13.47	15.90	16.33	15.99	**13.19**
SSIM	0.8075	0.9479	0.9346	0.9535	0.9500	0.9495	0.9536	**0.9564**
MS-SSIM	0.8965	0.9719	0.9636	0.9745	0.9721	0.9718	0.9741	**0.9757**
IW-SSIM	0.8673	0.9597	0.9473	0.9619	0.9587	0.9582	0.9617	**0.9636**
GMSD	0.1229	0.0441	0.0490	0.0434	0.0452	0.0467	0.0437	**0.0401**
FSIM	0.9439	0.9744	0.9716	0.9775	0.9764	0.9759	0.9782	**0.9795**
$FSIM_c$	0.9381	0.9737	0.9706	0.9767	0.9752	0.9746	0.9769	**0.9788**

Super-resolution		Training cost function			
Image quality metric	Bilinear	ℓ_2	ℓ_1	MS-SSIM	Mix
$1000 \cdot \ell_2$	2.5697	1.2407	1.1062	1.3223	**1.0990**
PSNR	27.16	30.66	31.26	30.11	**31.34**
$1000 \cdot \ell_1$	28.7764	20.4730	19.0643	22.3968	**18.8983**
SSIM	0.8632	0.9274	0.9322	0.9290	**0.9334**
MS-SSIM	0.9603	0.9816	0.9826	0.9817	**0.9829**
IW-SSIM	0.9532	0.9868	0.9879	0.9866	**0.9881**
GMSD	0.0714	0.0298	0.0259	0.0316	**0.0255**
FSIM	0.9070	0.9600	0.9671	0.9601	**0.9680**
$FSIM_c$	0.9064	0.9596	0.9667	0.9597	**0.9677**

JPEG de-blocking		Training cost function			
Image quality metric	Original JPEG	ℓ_2	ℓ_1	MS-SSIM	Mix
$1000 \cdot \ell_2$	0.6463	0.6511	0.6027	1.9262	**0.5580**
PSNR	32.60	32.73	32.96	27.66	**33.25**
$1000 \cdot \ell_1$	16.5129	16.2633	16.0687	33.6134	**15.5489**
SSIM	0.9410	0.9427	0.9467	0.9364	**0.9501**
MS-SSIM	0.9672	0.9692	0.9714	0.9674	**0.9734**
IW-SSIM	0.9527	0.9562	0.9591	0.9550	**0.9625**
GMSD	0.0467	0.0427	0.0413	0.0468	**0.0402**
FSIM	0.9805	0.9803	0.9825	0.9789	**0.9830**
$FSIM_c$	0.9791	0.9790	0.9809	0.9705	**0.9815**



(a) Clean image (b) Clean (c) Noisy (d) ℓ_2 (e) ℓ_1 (f) SSIM (g) MS-SSIM (h) Mix (i) Clean (j) Noisy (k) ℓ_2 (l) ℓ_1 (m) SSIM (n) MS-SSIM (o) Mix

Fig. 1: Comparisons of the results of joint denoising and demosaicking performed by networks trained on different loss functions (best viewed in the electronic version by zooming in). ℓ_2, the standard loss function for neural networks for image processing, produces splotchy artifacts in flat regions (d).

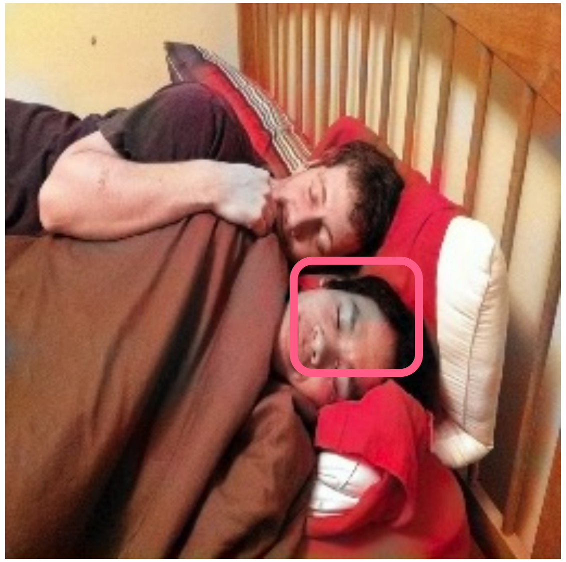
* https://arxiv.org/abs/1511.08861
* https://medium.com/srm-mic/all-about-structural-similarity-index-ssim-theory-code-in-pytorch-6551b455541e
* https://ieeexplore.ieee.org/document/1292216,

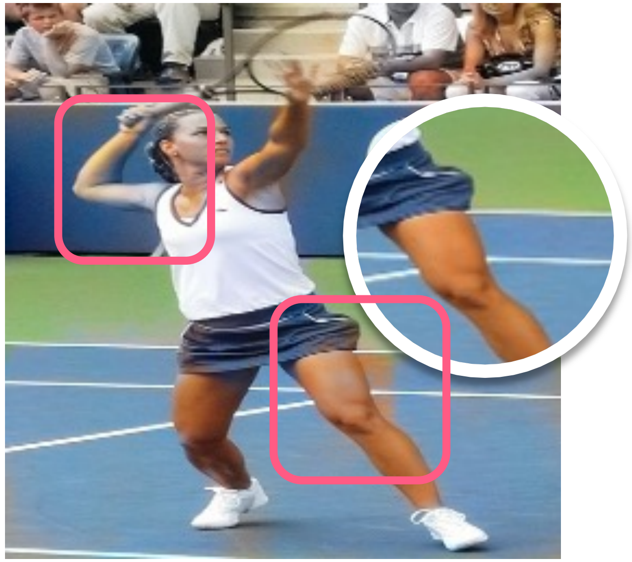
Ⅳ. Color Bleeding Artifact

Residual Attention U-Net with SSIM Loss

· Color bleeding artifact occurs when segmentation is not performed properly.

· When Residual Attention U-Net was trained, it showed poor results with respect to the color bleeding artifact.





Ⅴ. Result

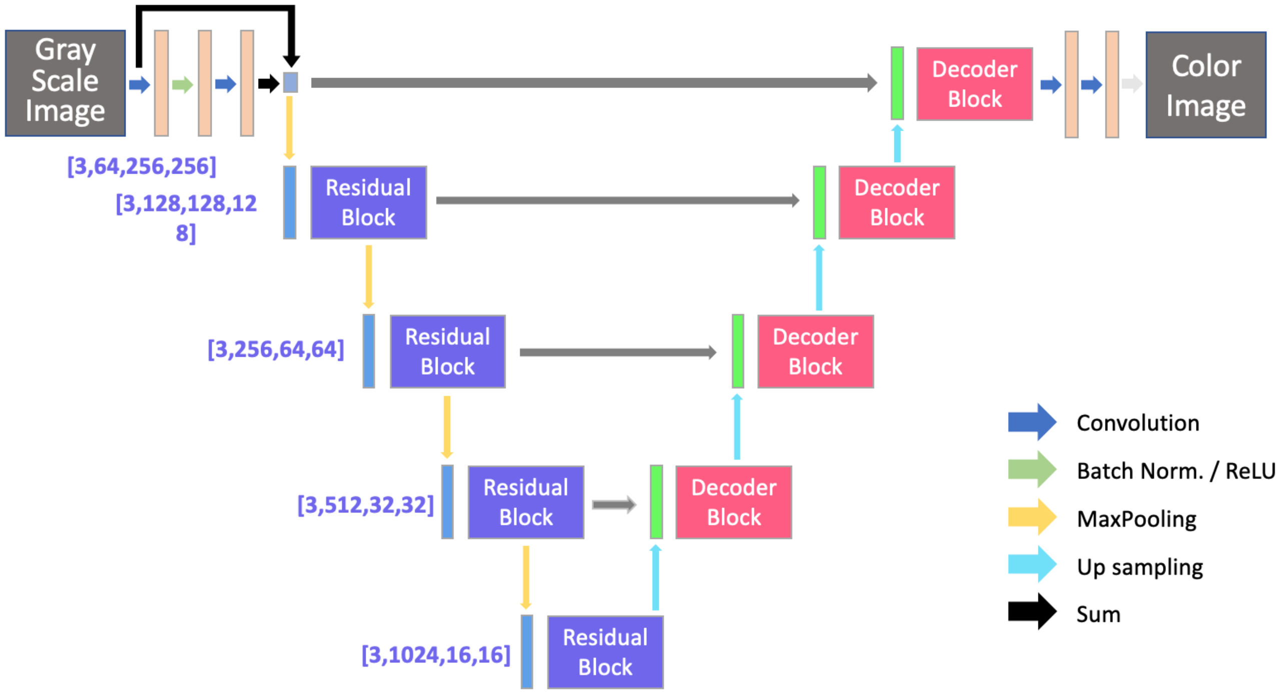
V. Best Model Information

Model Overview

- Residual U-Net
- 120 epochs
- L1 + MS SSIM Loss (α = 0.025)
- AdamW Optimizer
- Batch Size 4
- LR 2.5e-5

- Down Scaling with MaxPooling
- Up Scaling with Upsampling

- Contributed by Jeon Tak



V. Result

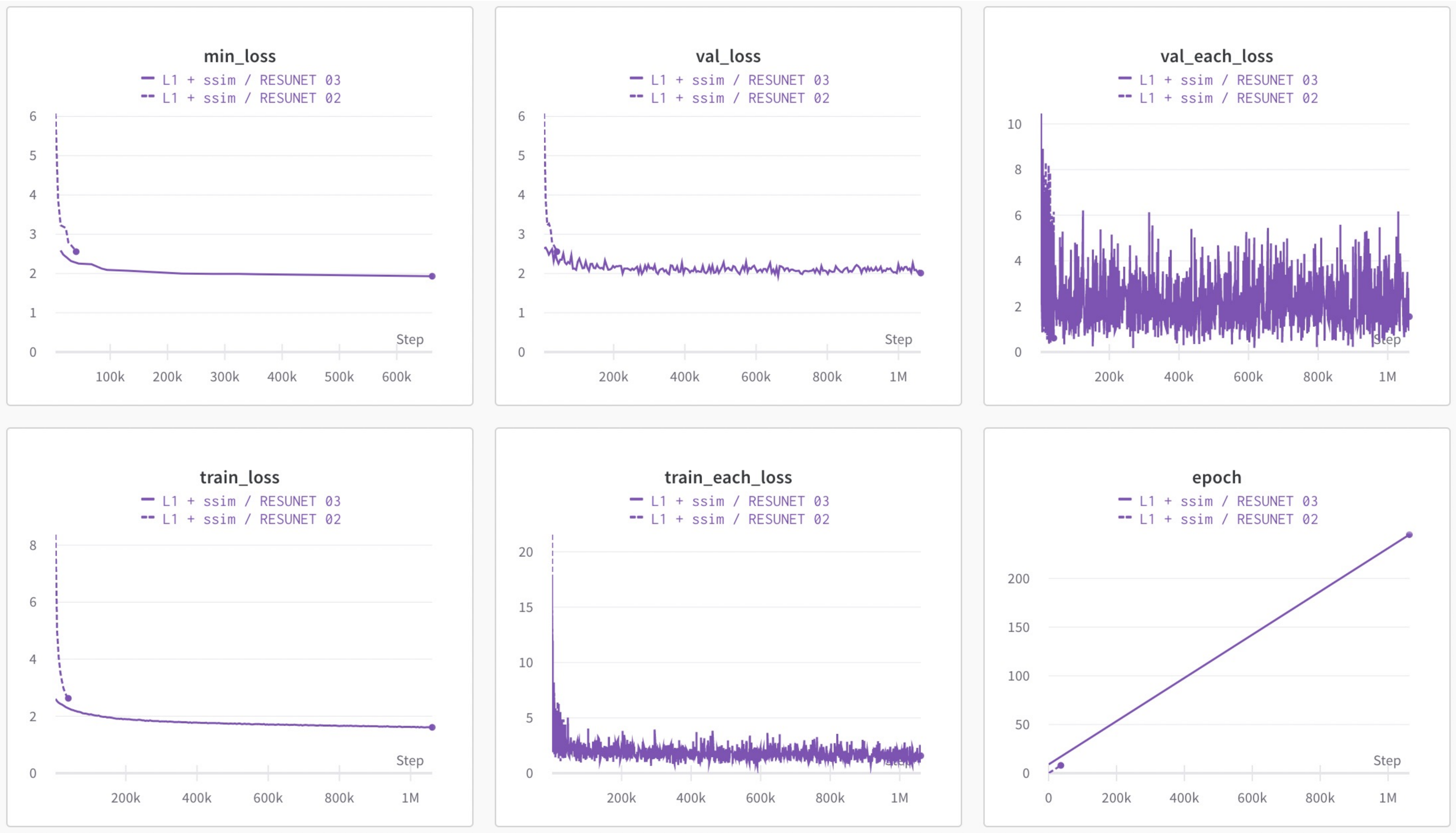
Ground Truth

Input Image(GrayScale)

Output Image(Color)

V. Result

V. Result

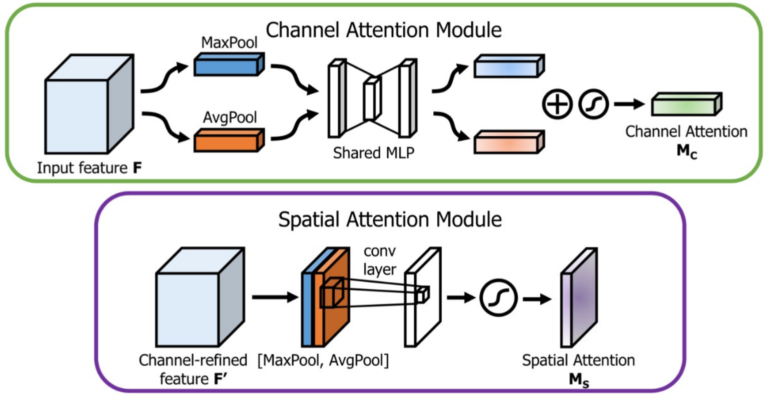


* Validation Loss: 2.085
* Training Loss: 1.739

Ⅴ. Result

Improvements

- Hyperparameter Tunning
  (L1+MS SSIM Loss α value, etc.)
- Data Augmentation
- Increase batch size (more than 4)
- Understanding of Attention Block
  - Channel Attention
  - Spatial Attention



Ⅴ. Result

Github Organization

https://github.com/GC221ComputerVision

Term Project Final Presentation

Computer Vision

| 2022.06.07 | Team 1

14446_001, Tue/Wed
Department of software

201533631 김도균
201735900 전탁
201935017 김민경
201935048 박준이
201935034 김한결
201935135 조수현