Graduation Study | Lab Seminar | Midterm Presentation

画像認識タスクに汎用な熱赤外線画像着色モデルの検討

General-Purpose Infrared Image Coloring Model for Various Recognition Tasks

Your feedback is welcome!

大町・宮崎研究室
学部4年 谷内寛人

Hiroto TANIUCHI, 4th year undergraduate,
Laboratory for Image Information Communications (IICLab),
Tohoku University

Outline

1. Introduction
2. Ugawa’s Model
3. Possibility and Motivation
4. Approach and Proposal
5. Experiment Design
6. Conclusion
7. Topic

1. Introduction

“Color” is interesting as research topic!
...Thermal infrared cameras are robust sensor that can withstand environmental changes.

Feature

Capture **the heat** of objects

- ☑No lighting required
- ☑Available in bad weather

Visible light image

https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/det_result.png

Thermal Infrared (**TIR**) Image

Applications

Automatic operation[1]

https://www.flir.com/globalassets/news/1200x628_autonomousvehicle.jpg

Rescue[1]

https://www.flir.jp/globalassets/defense/solution-and-landing-pages/uis/search-rescue-benner.jpg/constrain-1130x0-20105106 68.jpg

Crime prevention

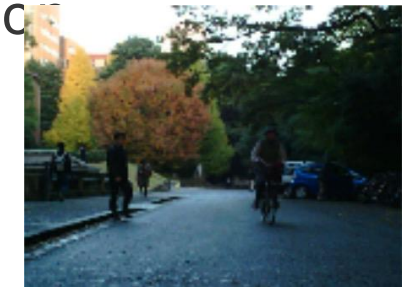
https://shopping.wtw.jp/cdn/shop/files/00000000776_kZHOT5k.png?v=1687514776

1. Introduction

Visible Light

❌ Influenced by the environment

✅ Easy to grasp the situation



https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/det_result.png

Thermal InfraRed

✅ Robust to changes in the environment

❌ Low visibility

https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/det_result.png

Color Texture

Heat Information

Fake Visible Light Image Generated from TIR Image

✅ Easy to grasp the situation

✅ Robust to environment

2. Ugawa's Model

- Proposed by Kei Ugawa, agraduate of IIC Lab
- TICC-GAN[3] used as baseline
- Generate colored images that properly reflect meaning of objects
- Refer to feature maps from the segmentation module

～2023/11/25
Memory of Ekiden～

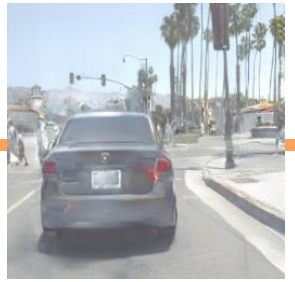
2. Ugawa's Model

Previous Coloring Model Using GAN

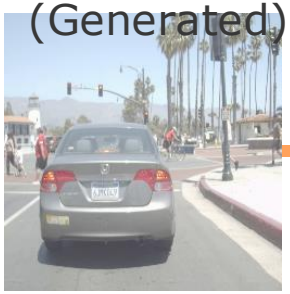
TIR Image I
(Original)

$\begin{cases} I \ldots Infrared\ ray \\ V \ldots Visible\ right \end{cases}$

Generator



Fake Visible Light Image $\hat{V}$
(Generated)

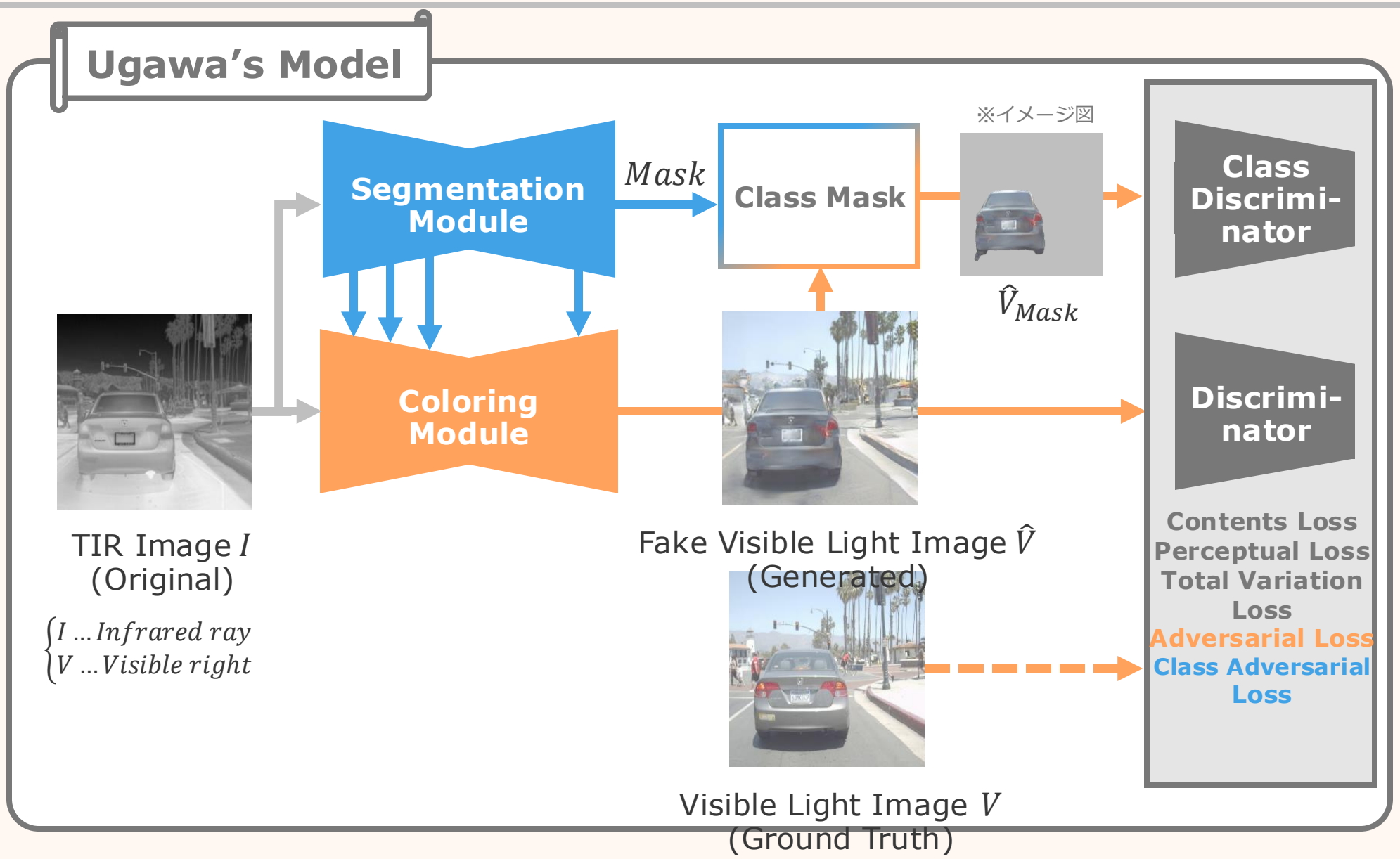


Visible Light Image V
(Ground Truth)

Discrimi-
nator

Contents Loss
Perceptual Loss
Total Variation Loss
Adversarial Loss

2. Ugawa's Model



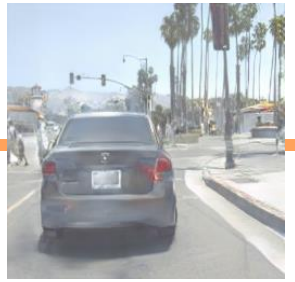
2. Ugawa's Model

Previous Coloring Model Using GAN

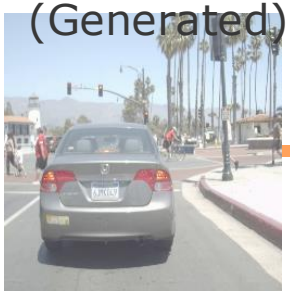
TIR Image I
(Original)

$\begin{cases} I \ldots Infrared\ ray \\ V \ldots Visible\ right \end{cases}$

Generator



Fake Visible Light Image $\hat{V}$
(Generated)



Visible Light Image V
(Ground Truth)

Discrimi-
nator

Contents Loss
Perceptual Loss
Total Variation
Loss
Adversarial Loss

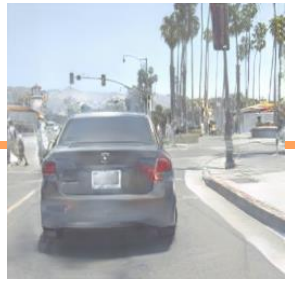
2. Ugawa's Model

Previous Coloring Model Using GAN -Loss-

TIR Image I
(Original)

$\begin{cases} I \ldots Infrared\ ray \\ V \ldots Visible\ right \end{cases}$

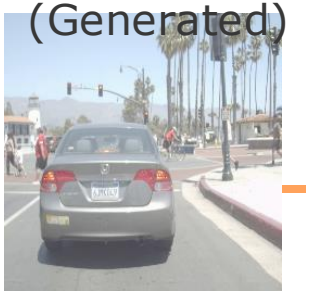
Generator



Fake Visible Light Image $\hat{V}$
(Generated)

Discrimi-
nator

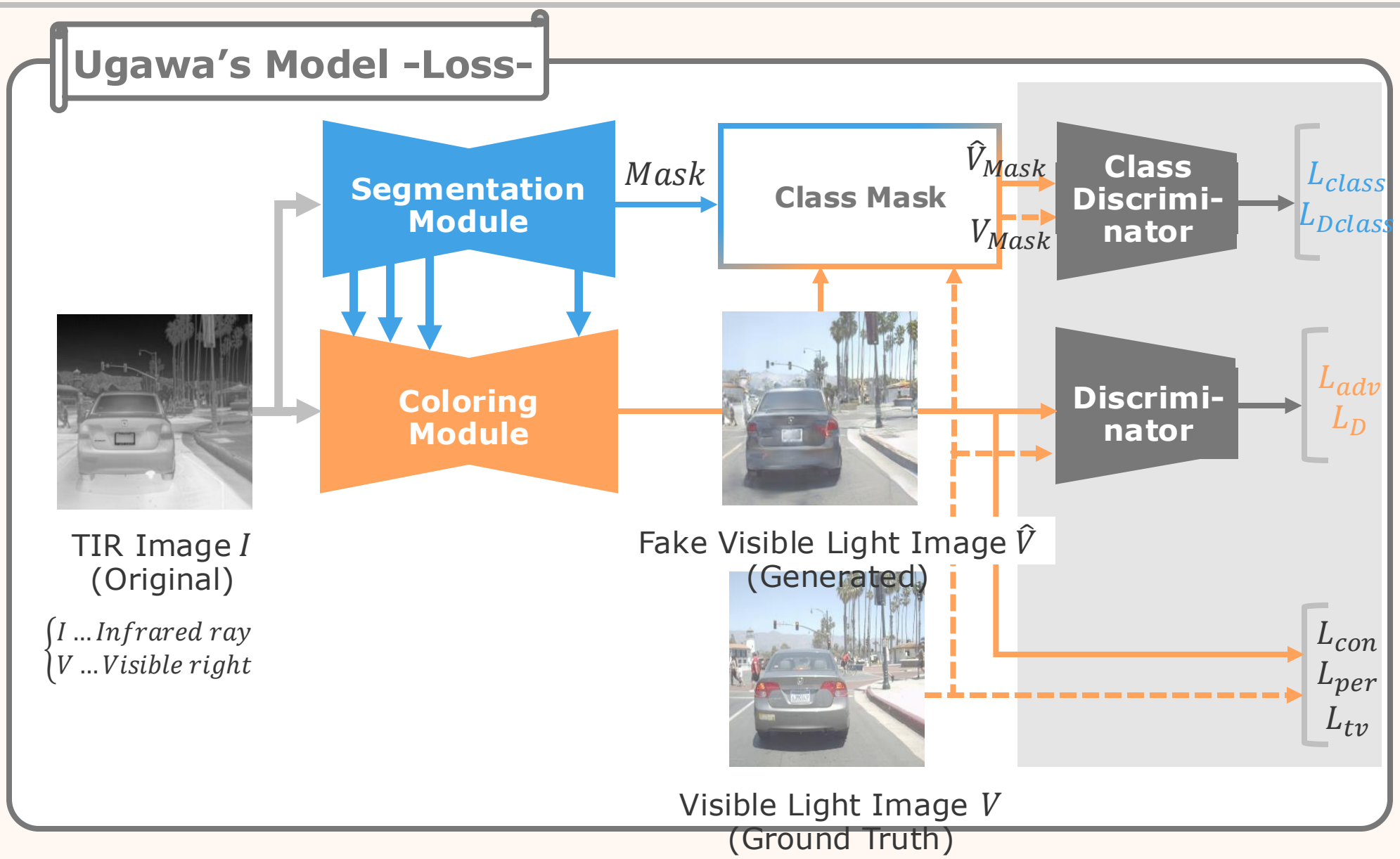
L_{adv}
L_D



Visible Light Image V
(Ground Truth)

L_{con}
L_{per}
L_{tv}

2. Ugawa's Model



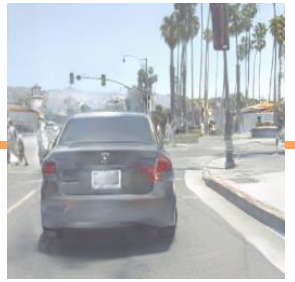
2. Ugawa's Model

Previous Coloring Model Using GAN -Loss-

TIR Image I
(Original)

$\begin{cases} I \ldots Infrared\ ray \\ V \ldots Visible\ right \end{cases}$

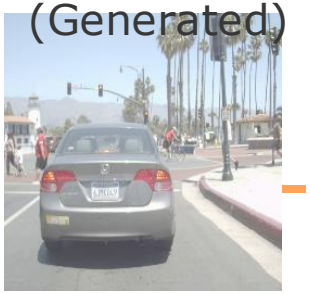
Generator



Fake Visible Light Image $\hat{V}$
(Generated)

Discrimi-
nator

L_{adv}
L_D



Visible Light Image V
(Ground Truth)

L_{con}
L_{per}
L_{tv}

3. Possibility and Motivation

Viewpoint

- As part of the experiment, input the fake visible right image $\hat{V}$ to segmentation model $Model^S$ **trained by visible light images**.
- Lack of coloring accuracy and could not classify accurately. (Originally the goal of the model is to improve image quality.)
- If the output of the model works when we input it to a model **trained by visible light images,** it means we can apply TIR images to Publicly available trained large-scale models! ex) ResNet, BiT, CLIP...
- In short, we would like to covert TIR image while taking account naturalness not only **for human**, but also **for recognition model**.

Motivation

Using output image $\hat{V}$ from Ugawa's model as an input for recognition models trained by visible light images

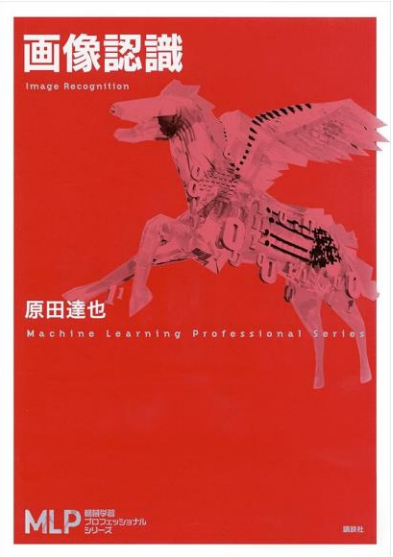
4. Approach and Proposal

Dataset should have visible light image, TIR image and annotation aligned on the same scene.

MFNet Dataset

- Natural images of parking and roads
- Night and day
- 640x480 pixels
- total 1606
- Made in University of Tokyo [4]

https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/predictionExamples_good.png



https://m.media-amazon.com/images/I/91TCHCgvJ2L._SY425_.jpg

4. Approach and Proposal

Proposed Method

MFNet データセット

TIR Image I
(Original)

$\begin{cases} I \ldots Infrared\ ray \\ V \ldots Visible\ right \\ S \ldots Segmentation \end{cases}$

Segmentation Module

$Mask$

Class Mask

※イメージ図

$\hat{V}_{Mask}$

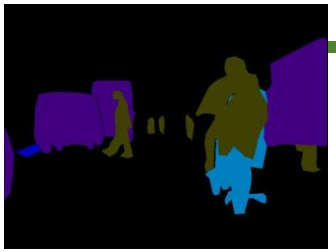
Coloring Module

Fake Visible Light Image $\hat{V}$
(Generated)

Segmentation Model Trained by Visible Light

$\hat{S}$

Visible Light Image V
(GT①)



Segmentaion S
(GT②)

Class Discrimi-nator

Discrimi-nator

Contents Loss
Perceptual Loss
Total Variation Loss
Adversarial Loss
Class Adversarial Loss
Segmentation Loss

4. Approach and Proposal

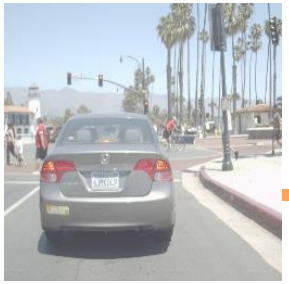
Previous Coloring Model Using GAN

TIR Image I
(Original)

$\begin{cases} I \ldots Infrared\ ray \\ V \ldots Visible\ right \end{cases}$

Coloring Module

Fake Visible Light Image $\hat{V}$
(Generated)

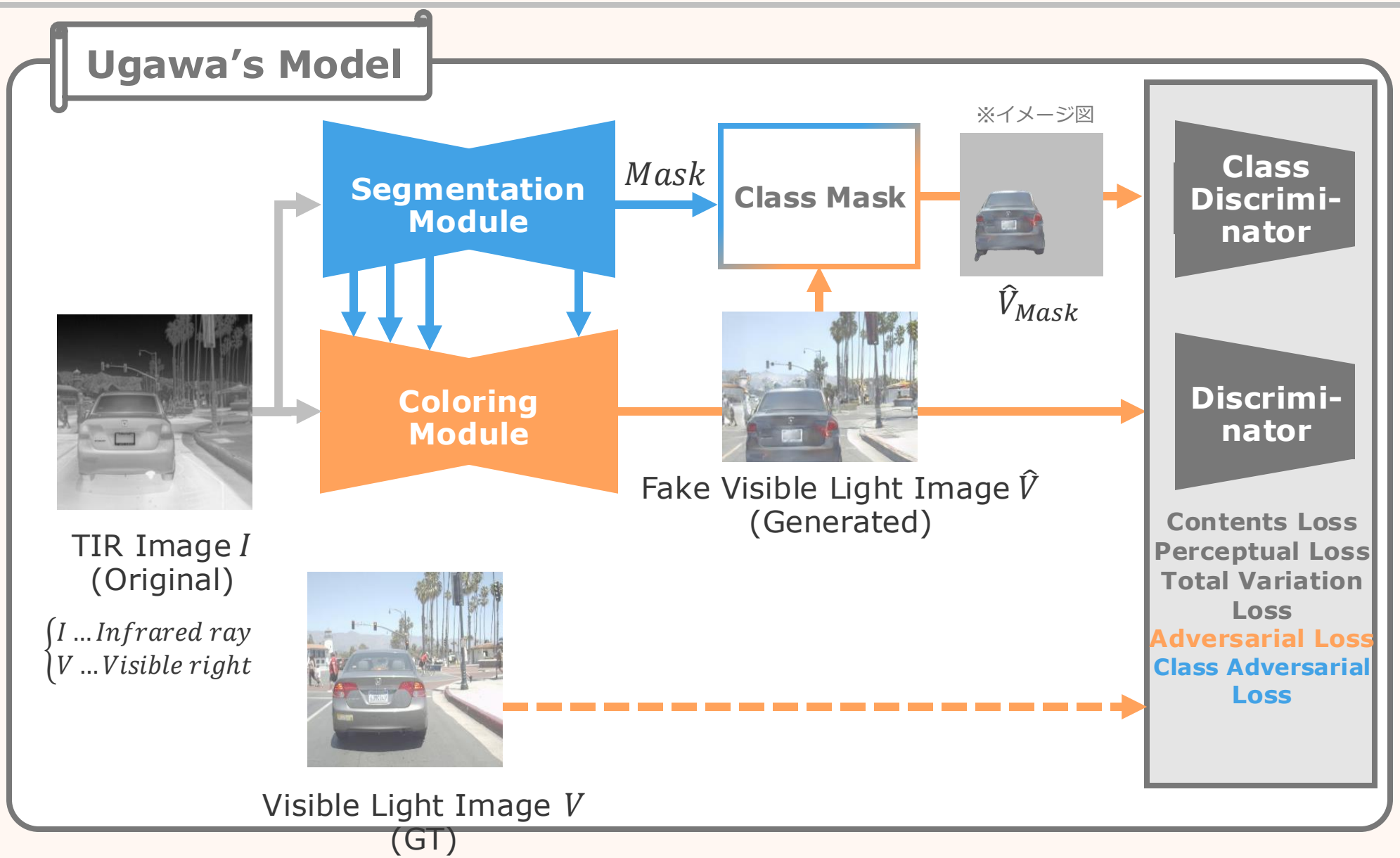


Visible Light Image V
(GT)

Discrimi-nator

Contents Loss
Perceptual Loss
Total Variation Loss
Adversarial Loss

4. Approach and Proposal



4. Approach and Proposal

Proposed Method

MFNet データセット

TIR Image I
(Original)

$\begin{cases} I \ldots Infrared\ ray \\ V \ldots Visible\ right \\ S \ldots Segmentation \end{cases}$

Segmentation Module

$Mask$

Class Mask

※イメージ図

$\hat{V}_{Mask}$

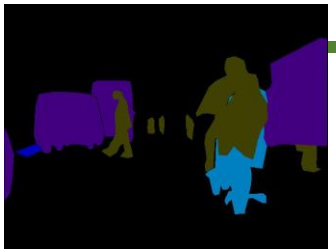
Coloring Module

Fake Visible Light Image $\hat{V}$
(Generated)

Segmentation Model Trained by Visible Light

$\hat{S}$

Visible Light Image V
(GT①)



Segmentaion S
(GT②)

Class Discrimi-nator

Discrimi-nator

Contents Loss
Perceptual Loss
Total Variation Loss
Adversarial Loss
Class Adversarial Loss
Segmentation Loss

5. Experiment Design

☑ ~~Run Ugawa's model stored on the server~~

☐ Follow-up experiment under the same conditions as those of Ugawa's paper, and obtain values close to the results by ourselves.

Midterm Presentation of Dec.

☐ (Implement the Ugawa's model on our own)

☐ Input the infrared images I of the MFNet dataset to the trained Ugawa's model $Model^U$ and obtain a fake light image $\hat{V}$

☐ Input $\hat{V}$ to the segmentation model $Model^S$ trained by visible light images and compare classification accuracy with real visible light version
= Confirm that the output $S^{\hat{V}}$ with $\hat{V}$ input is less accurate than the output S^V with V input.

End of Jan.

☐ Train $Model^{U'}$ adding "classification accuracy at $\hat{V}$ input" into loss
= Attempt to maintain "naturalness as seen by humans" and make it "work well even when input to a model pretrained with visible light images

Final Presentation of Feb.

6. Conclusion

Motivation

Using output image $\hat{V}$ from Ugawa's model as an input for recognition models trained by visible light images

Approach

- Add the gap between ground truth$S$ and "segmentated $\hat{V}$ " $\hat{S}$ (=unnaturalness for recognition model) to the loss function of the discriminator
- Introducing MFNet, a dataset with visible light image, TIR image and annotation aligned **on the same scene**
- Finally generalize to obgect detection tasks

7. Topic

- Is the novelty of this research theme recognized?
- I'm in pain from the error resolution. Crying.
- How to manage experimental conditions?
- How to connect and recombine existing models?
- How to verify what the image recognition model is looking at (Grad Cam?)
- How to explore repositories on GitHub?
- How to version control with GitHub? ...etc

Thank you for your attention!

Your feedback is welcome!

終
制作・著作

IIC

References

[1] **FLIR Systems** homepage https://www.flir.jp/

[2] **Kei Ugawa.** A Study on Thermal Infrared Image Colorization Based on Semantic Information, 2024.

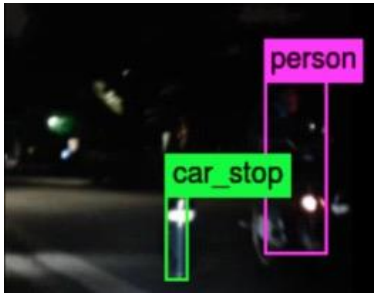
[3] **X. Kuang et al.** 'Thermal infrared colorization via conditional generative adversarial network', Infrared Physics & Technology, vol. 107, p. 103338, Jun. 2020, doi: 10.1016/j.infrared.2020.103338.

[4] **Qishen Ha, Kohei Watanabe, Takumi Karasawa, Yoshitaka Ushiku, Tatsuya Harada.** MFNet: Towards Real-Time Semantic Segmentation for Autonomous Vehicles with Multi-Spectral Scenes. The 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017), 2017. https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/

Prospects and Ideas for Master's Degree

Continue undergraduate research

- Generalization to other tasks
  = Use of **object detection** data contained in MFNet
- Lightweight Ugawa's Model



https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/det_result.png

Other themes related to color

- Evaluate the amount of semantic information that is lost when converted to monochrome.
- Quantitatively evaluate how easy it is for colorblind people to see the display of a package, etc.
- Implement a service that provides color schemes for slides according to the input of feelings and themes using words.

Feedback Obtained

- Possibly more effective using diffusion models than GAN

Supplementary Material–Ugawa's Model Generator Loss Functions-

$$L_{all} = L_{con} + \lambda_{adv} L_{adv} + L_{per} + L_{tv} + \underbrace{\lambda_{class} L_{class}}_{\text{損失の重み}}$$

- 既存手法[2]で用いられている4つ損失関数に加えて、クラス敵対的損失を追加したものを生成器の学習に用いる損失関数とする

- コンテンツ損失 L_{con} ： 生成画像とGround-truthの各画素の差分の絶対値（L1損失）
- 敵対的損失 L_{adv} ： 生成画像が判別器に(Ground-Truthよりも相対的に)本物らしいと判断されると小さくなる損失
- 知覚的損失 L_{per} ： 生成画像とGround-truthを学習済VGGに入力した際の特徴マップのL1損失
- Total Variation損失 L_{tv} ： 画像のx,y方向それぞれの勾配を損失とするもの
- **クラス敵対的損失 L_{class}** ： マスクをもとに特定クラスの領域を切り抜いた画像の敵対的損失