

学部卒業研究

研究室ゼミ

認識タスクに適した 熱赤外線画像 着色モデルに関する研究

Infrared Image Coloring Model
for Image Recognition Tasks

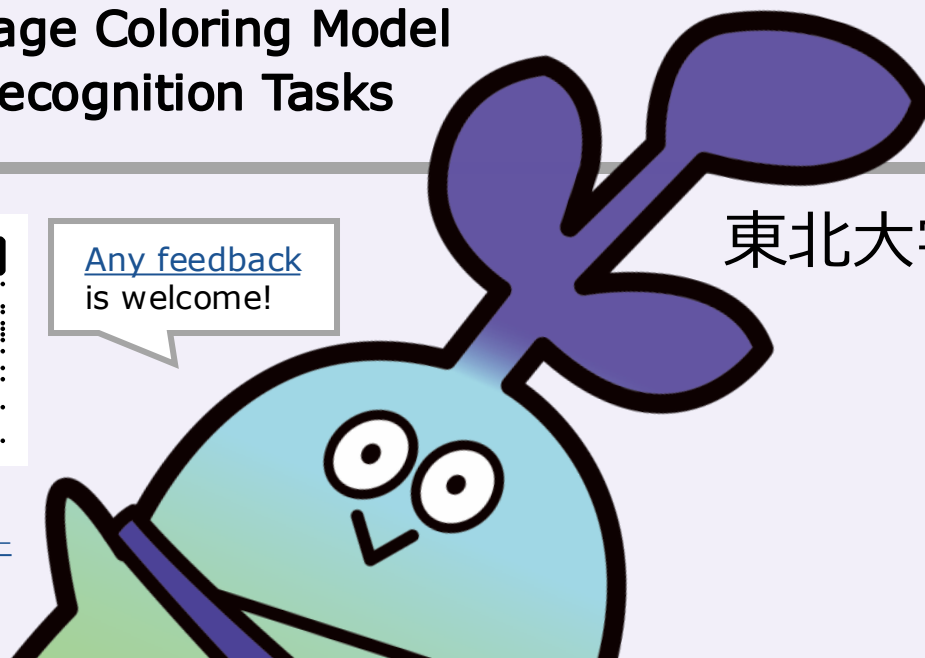


[Any feedback](#)
is welcome!

東北大学広報活動
キャラクター・研一

東北大学 大町宮崎研究室
学部4年 谷内寛人

Hiroto TANIUCHI, 4th year undergraduate,
[IICLab](#), Tohoku University



目次

1. 導入 —赤外線画像の着色—
2. 関連研究 —宇川モデル—
3. 宇川モデルの発展性, 研究の動機, 提案手法
4. 実験
5. 結果, 考察
6. 結論, 展望

Outline

1. Introduction
2. Ugawa's Model
3. Motivation, Approach, Proposal
4. Experiment
5. Result
6. Conclusion

1. 導入 ー赤外線画像の着色ー

“色”, 特に“色彩”について研究したい！
...環境変化に強い頑健なセンサとして熱赤外線カメラが注目されている

特徴

物体の熱を捉える

- ☑照明を必要としない = 夜でも使える
- ☑悪天候でも使用可能



可視光画像

https://www.m.t.u-tokyo.ac.jp/static/projects/mi_multispectral/det_result.png



熱赤外線画像

応用例



自動運転^[1]

https://www.flir.com/globalassets/news/120x628_autonomousvehicle.jpg



救助活動^[1]

<https://www.flir.jp/globalassets/defense/solution-and-landing-pages/ui5/search-rescue-bencher.jpg/constrain-1130x0-2010510668.jpg>



防犯

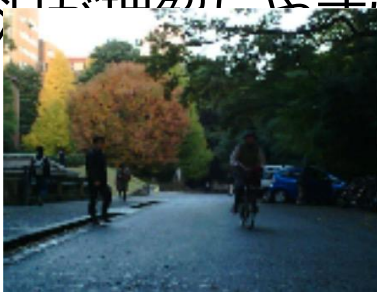
https://shop.ping.wtw.jp/cdn/shop/files/00000000776_k_ZH0T5K.png?v=1687514776

1. 導入 ー赤外線画像の着色ー

可視光..RGB画像 (Visible Light)

✗ 環境の影響を強く受ける

✓ 状況が理解しやすい



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

色情報テクスチャ

熱赤外線..TIR画像 (Thermal InfraRed)

✓ 天候・照明の変化に頑健

✗ 視認性が低く状況理解が困難



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

熱の画像情報

TIR画像から生成した擬似RGB画像

✓ 状況が理解しやすい ✓ 天候・照明の変化に頑健 ...と期待される

2. 関連研究 一宇川モデルー

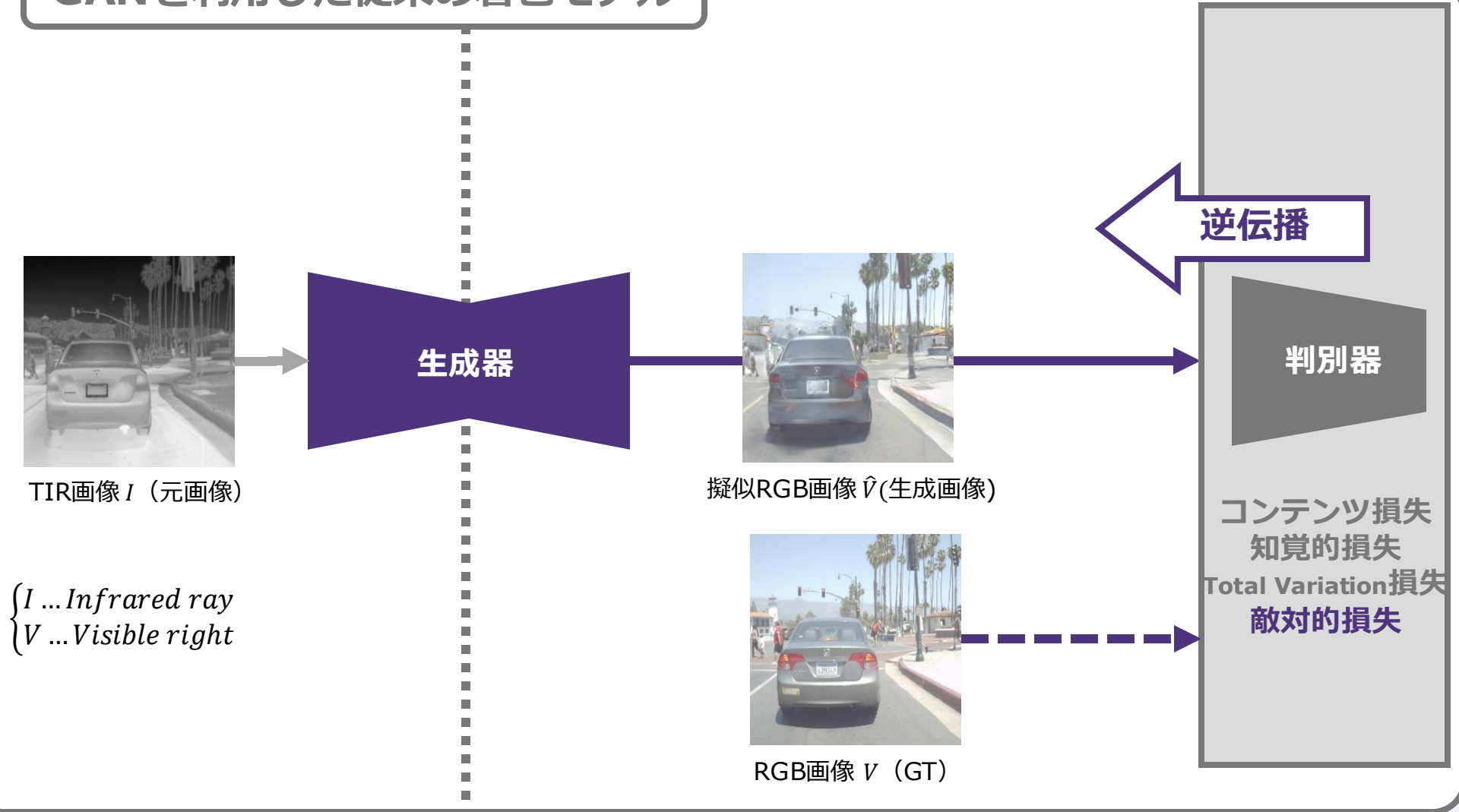
- ・ 大町・宮崎研究室2023年度卒業生の宇川さんが提案したモデル
- ・ TICC-GAN^[3](敵対的生成ネットワークの一つ)をベースラインに使用
- ・ セグメンテーションモジュールからの特徴マップを参照することで**物体の意味を適切に反映**した着色画像を生成



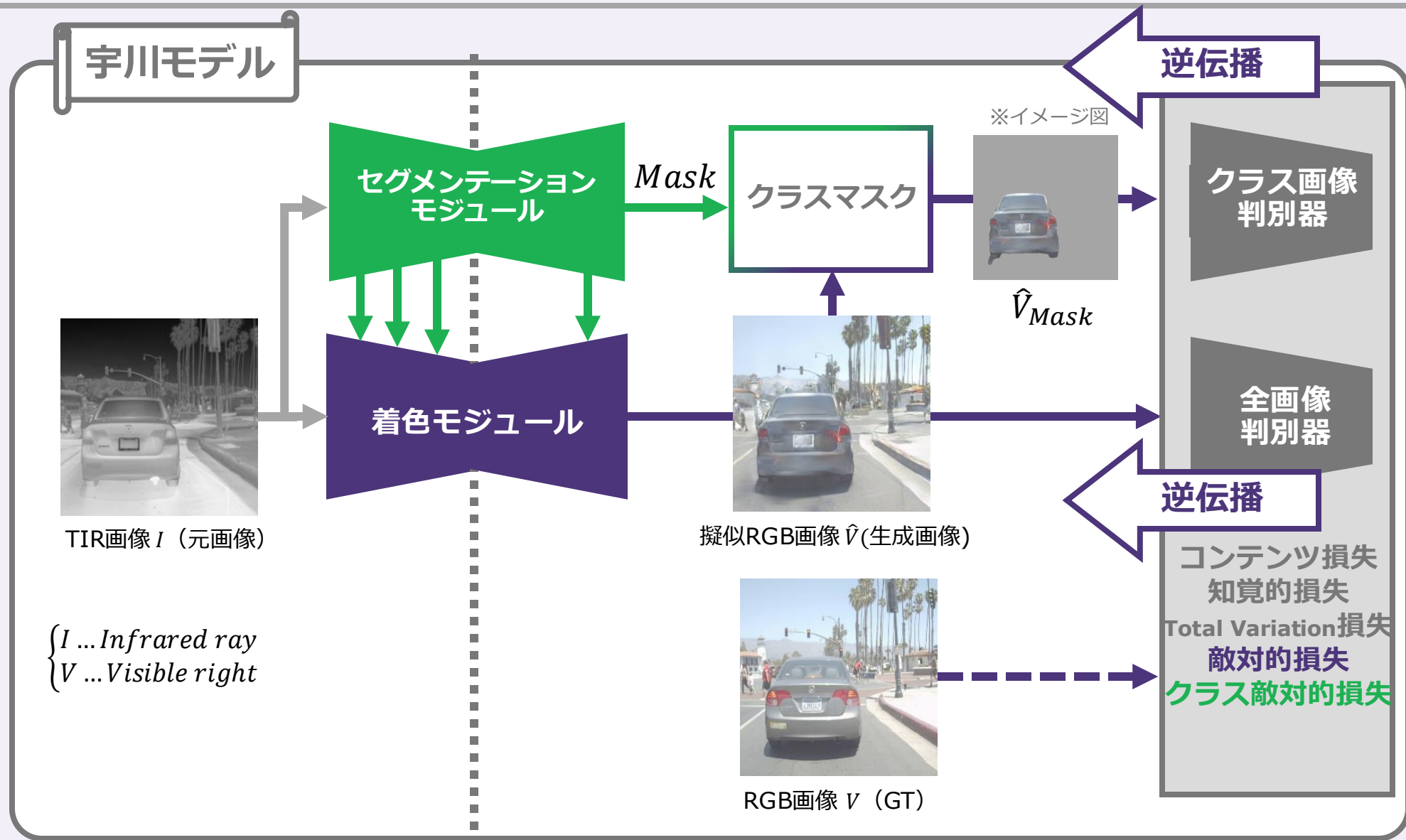
～2023/11/25
駅伝のおもひで～

2. 関連研究 一宇川モデル

GANを利用した従来の着色モデル



2. 関連研究 一宇川モデル



3. 宇川モデルの発展性, 研究の動機

着眼点

宇川モデルは認識モデルにとって自然な画像を生成できない

実験, 考察

- ・ 擬似RGB画像 $\hat{V}$ をRGB画像で学習済セグメンテーションモデル $Model^S$ に入力。分類性能は高くなかった。
- ・ もし、RGB画像で学習済のモデルに入力してもうまく動作するならば、一般公開されているパラメタをTIR画像に適用できる
- ・ すなわち“人にとって”だけでなく“画像認識モデルにとって”の自然さも考慮しながら変換を行いたい。

動機

着色結果 $\hat{V}$ を, RGB画像で訓練済みの認識モデル (セグメンテーション, 物体検出etc) の入力に利用したい



宇川モデル



大規模認識モデル
(セグメンテーション,
物体認識)

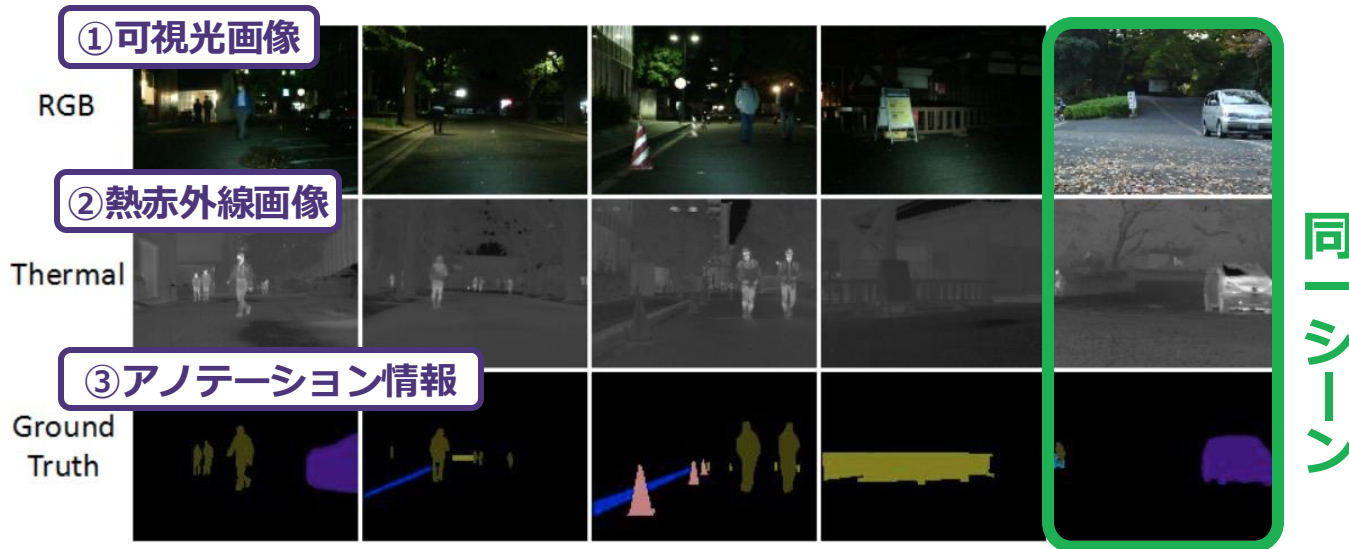


3. 宇川モデルの発展性, 研究の動機

①可視光画像②熱赤外線画像③アノテーション情報が同一シーン上で揃ったデータセットが必要

MFNet データセット

- ・ 街路や交通の画像 ・ 画素数 H480×W640
- ・ 昼間 : train 410枚, val 205枚, test 205枚
夜間 : train 374枚, val 187枚, test 188枚 } 計1542枚/全1606枚
- ・ 8クラス分類 (car, person, bike, curve, car stop, guardrail, color cone, bump, 無分類)



4. 実験方法

TIR⇒RGBのモデル

宇川モデル

- MFNet_TIR (昼間)→MFNet_RGB(昼間)を学習
- 宇川モデルのセグメンテーションモジュールは、FLIR及びIRVIデータセットにて事前学習させる

RGB⇒Segのモデル

DeeplabV3

- MFNet_RGB(昼間)→MFNet_Seg(昼間)を学習
- 学習済パラメタを利用

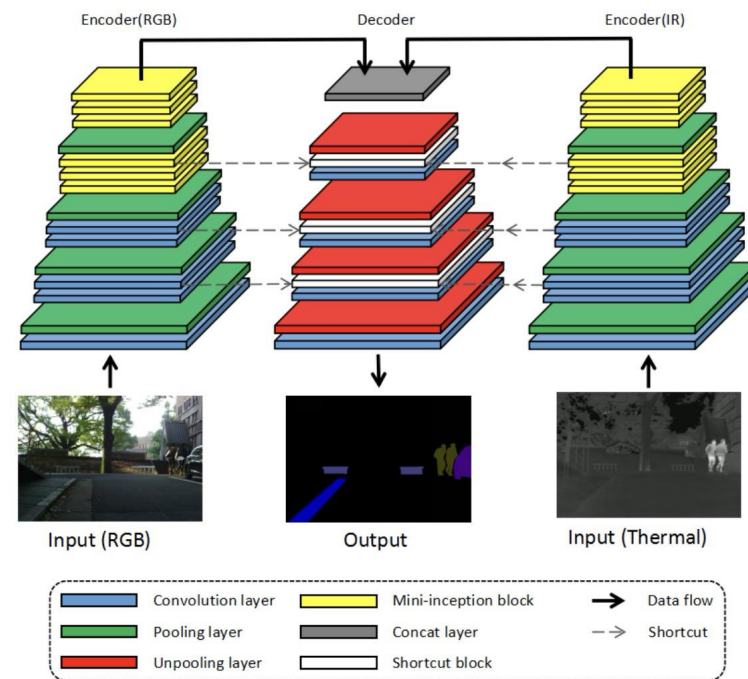
諸条件

- optimizer: SGD
- lr: 1e-3
- GPU: RTX A6000, GeForce RTX 3090

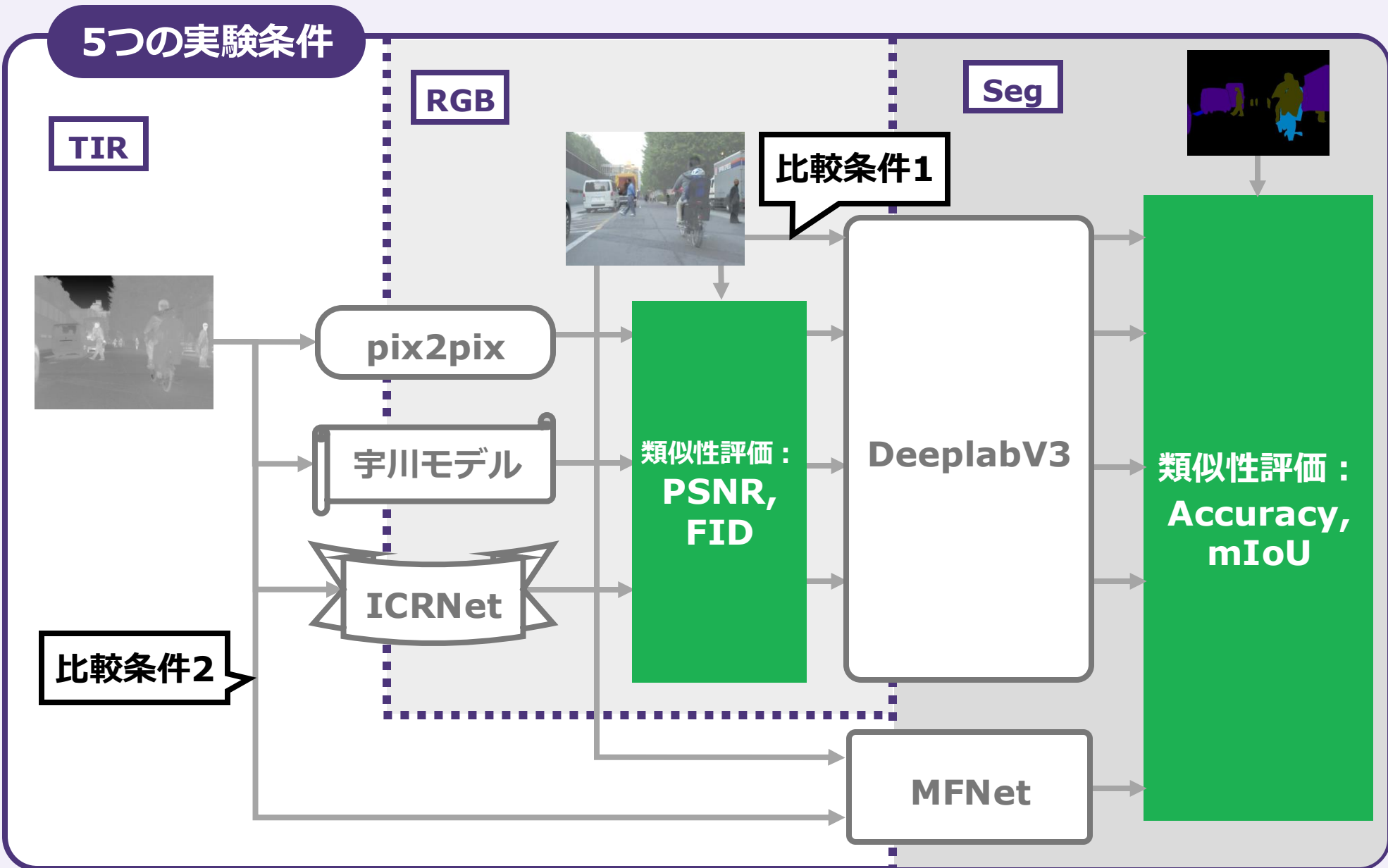
TIR&RGB⇒Segのモデル

MFNet (論文値参考)

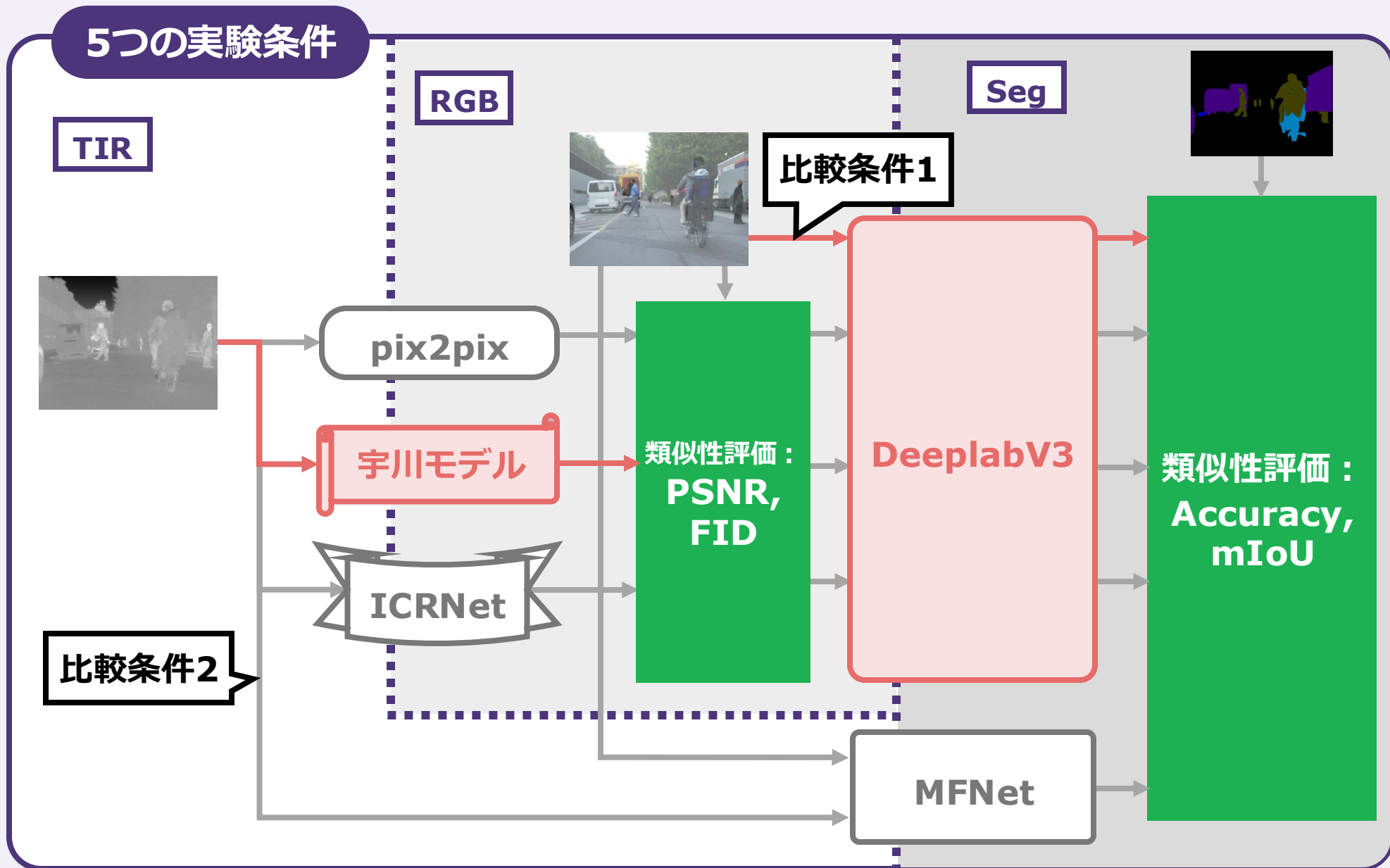
- 比較対象
- MFNet_RGB&TIR (昼)→MFNet_Seg(昼)を学習



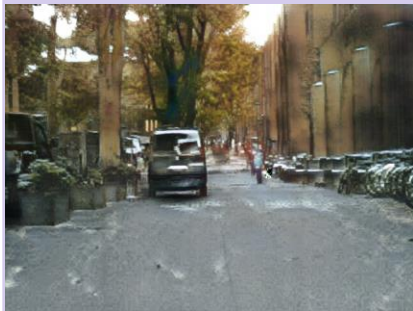
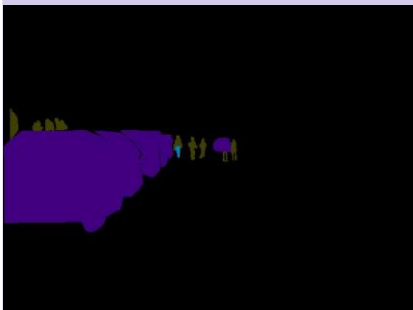
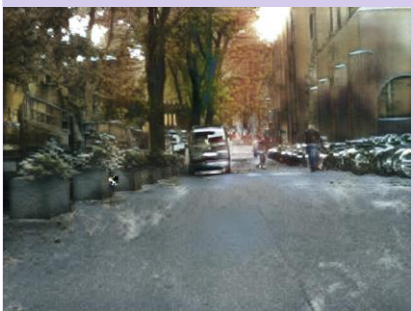
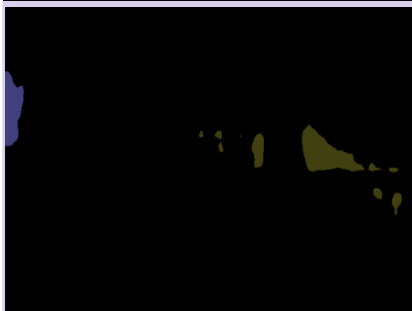
4. 実験方法



4. 実験方法



5. 定性評価

入力画像	Ground Truth	宇川モデル	Ground Truth	DeeplabV3
TIR	RGB	(TIR⇒)RGB	Seg	(RGB⇒)Seg
				
				
				

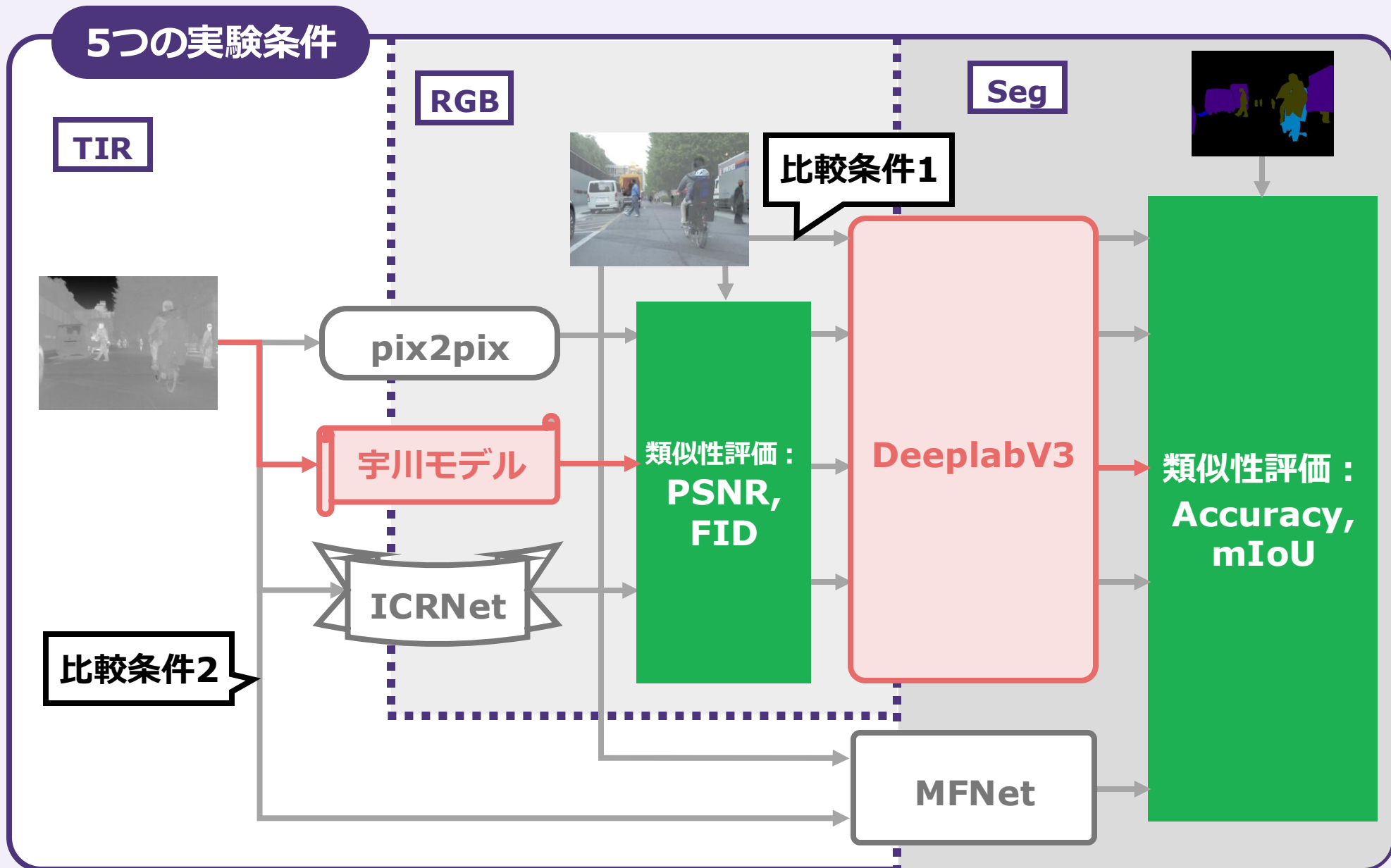
5. 定量評価

着色モデル	pix2pix	宇川モデル	ICRNet
ドメイン	TIR ⇨ RGB	TIR ⇨ RGB	TIR ⇨ RGB
評価指標			
PSNR (↑)		28.3	
FID (↓)		158.6	

着色性能 ↔ トレードオフ? ↔ 認識可能性

				比較条件1	比較条件2
セグモデル	DeeplabV3				MFNet
ドメイン	TIR⇨RGB	TIR⇨RGB	TIR⇨RGB	RGB ⇨ Seg	RGB&TIR⇨Seg
評価指標	→Seg	→Seg	→Seg		
Accuracy (↑)				0.184	0.591(参考) ^[4]
mIoU (↑)				0.145	0.649(参考) ^[4]

4. 実験方法



5. 考察

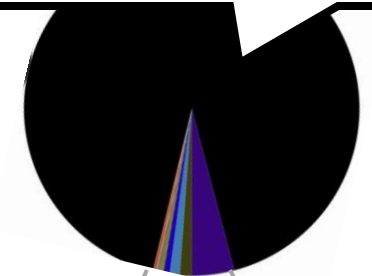
DeeplabV3の分類性能が極端に低い

- ①MFNetはTIR情報をconcatした4チャンネルの画像を入力とする=より豊富な情報を参照できる=有利
- ②MFNetデータセットはクラス情報が不均衡
 - ・総画素数のうち無分類の画素は9割以上、分類済クラスの中でもCarが半分以上を占める
 - ・不均衡に強い損失を採用してaccuracyを改善した

クラス 損失関数	0	1	2	3	4	5	6	7	8
CrossEntropy	0.91	0.52	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C.B. + Focul + Dice	0.91	0.75	0.0	0.0	0.0	0.0	0.0	0.0	0.0
(SegNet ^[4])(参考)	0.97	0.89	0.77	0.77	0.43	0.32	0.0	0.0	0.0
(MFNet ^[4])(参考)	0.97	0.56	0.91	0.75	0.77	0.33	0.0	0.72	0.71

各クラスの画素単位 出現頻度合計

Unlabelled(class0):
276868334 pixel=92.2%



Car(class1):
12315474 pixel=4.1%



5. 考察

分類性能が低地に低い

①M
入力

②M

・ 総
分

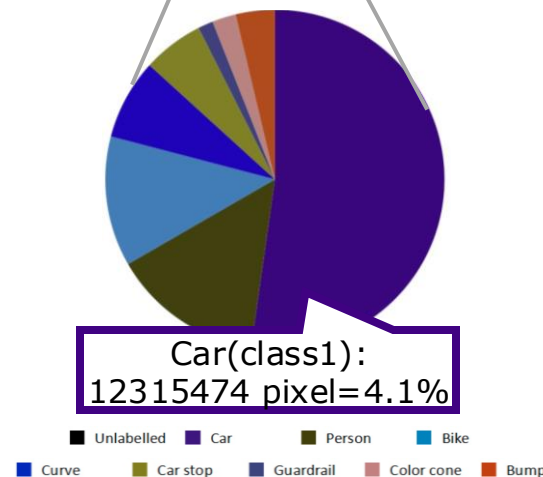
・ 不均衡に強い損失を採用して accuracy を改善

それにしても値が悪すぎる
→実装の際,参照先が間違っていた
可能性もある？

クラス 損失関数	0	1	2	3	4	5	6	7	8
CrossEntropy	0.91	0.52	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C.B. + Focul + Dice	0.91	0.75	0.0	0.0	0.0	0.0	0.0	0.0	0.0
(SegNet ^[4])(参考)	0.97	0.89	0.77	0.77	0.43	0.32	0.0	0.0	0.0
(MFNet ^[4])(参考)	0.97	0.56	0.91	0.75	0.77	0.33	0.0	0.72	0.71

各クラスの画素単位
頻度合計

Unlabelled(class0):
12315474 pixel=92.2%



5. 結論

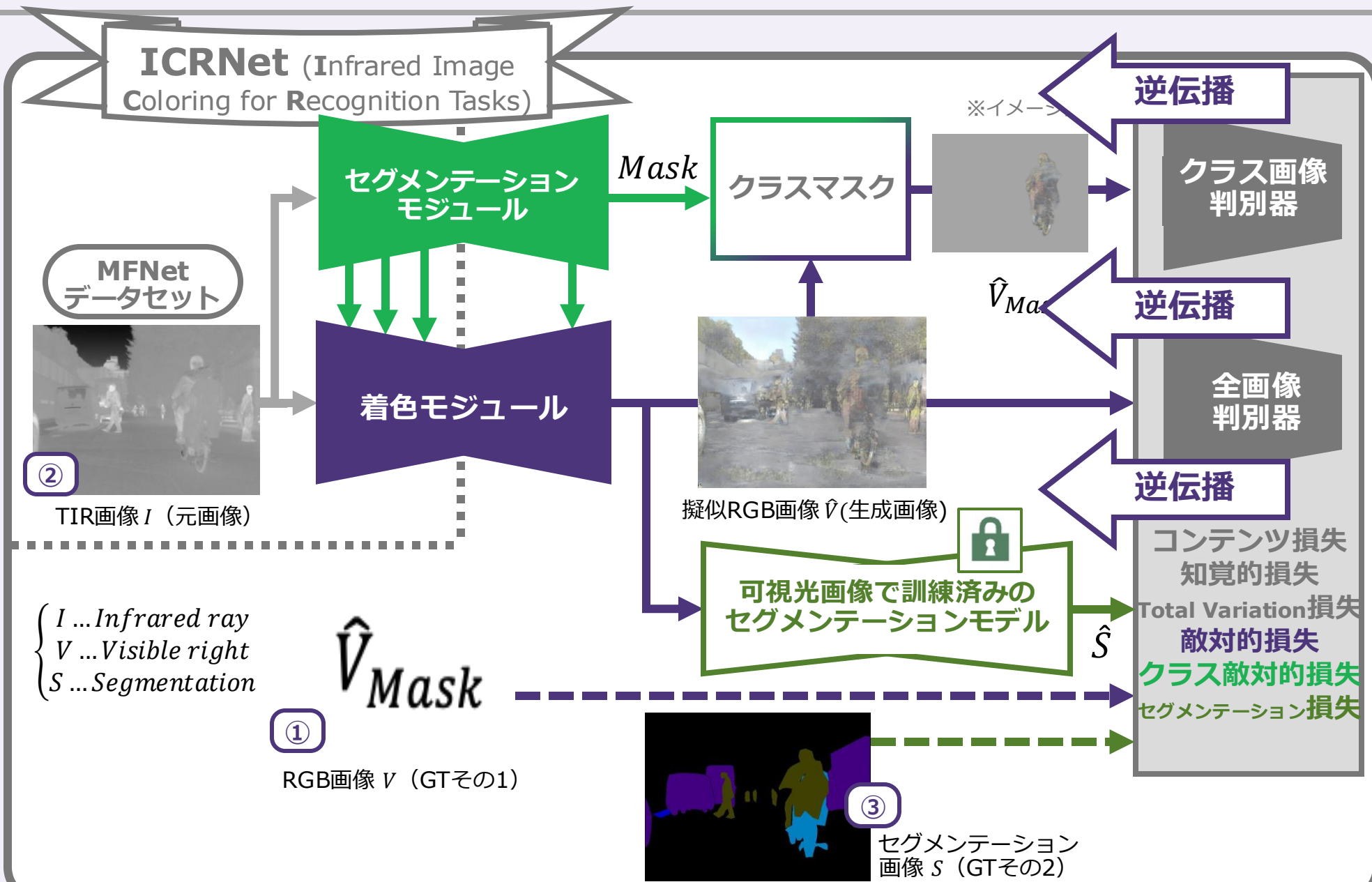
貢献

- ・ TIR, RGB, Segの3ドメインが同一シーン上に揃っているデータセットに宇川モデルを適用し、画像生成性能の高さを確認した
- ・ データセットの分布不均衡に強い損失関数を適用し、認識性能を向上した

展望

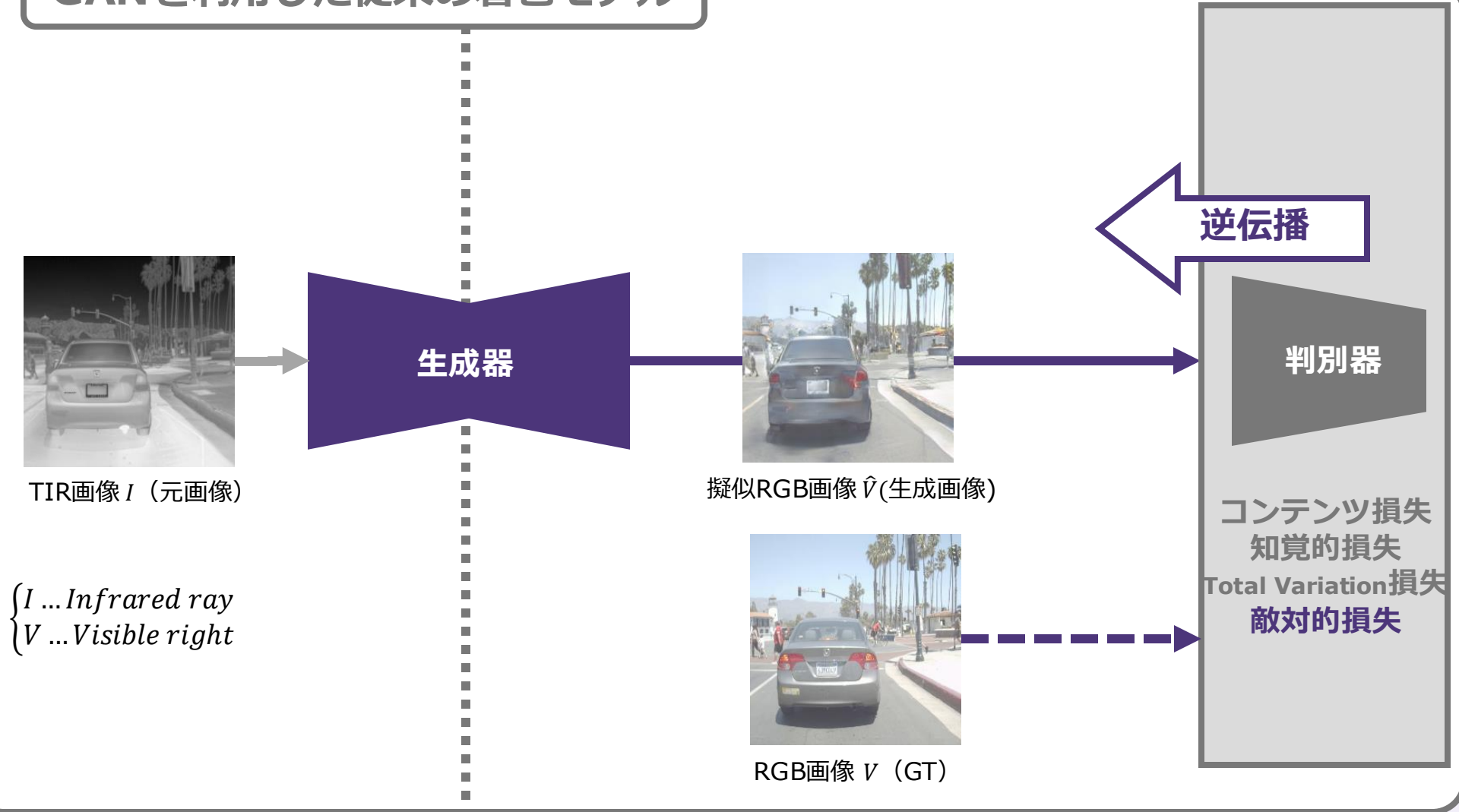
- ・ 学習元データセットの不均衡は分類性能に大きな影響を与える
 - 前景物体が大きく捉えられたデータセット（三ドメイン一体）を探す
 - 交通状況の把握という目的を踏まえ、MFNetの認識性能を向上させる
 - 画像データセットを自作する
- ・ DeeplabV3にも4チャンネル画像を訓練させてみる
 - 4チャンネル目(TIR)を宇川モデルに生成させるアーキテクチャの提案
- ・ MFNetには物体検出データも格納されている
 - セグメンテーション以外の認識タスクへの汎化

3. 提案手法

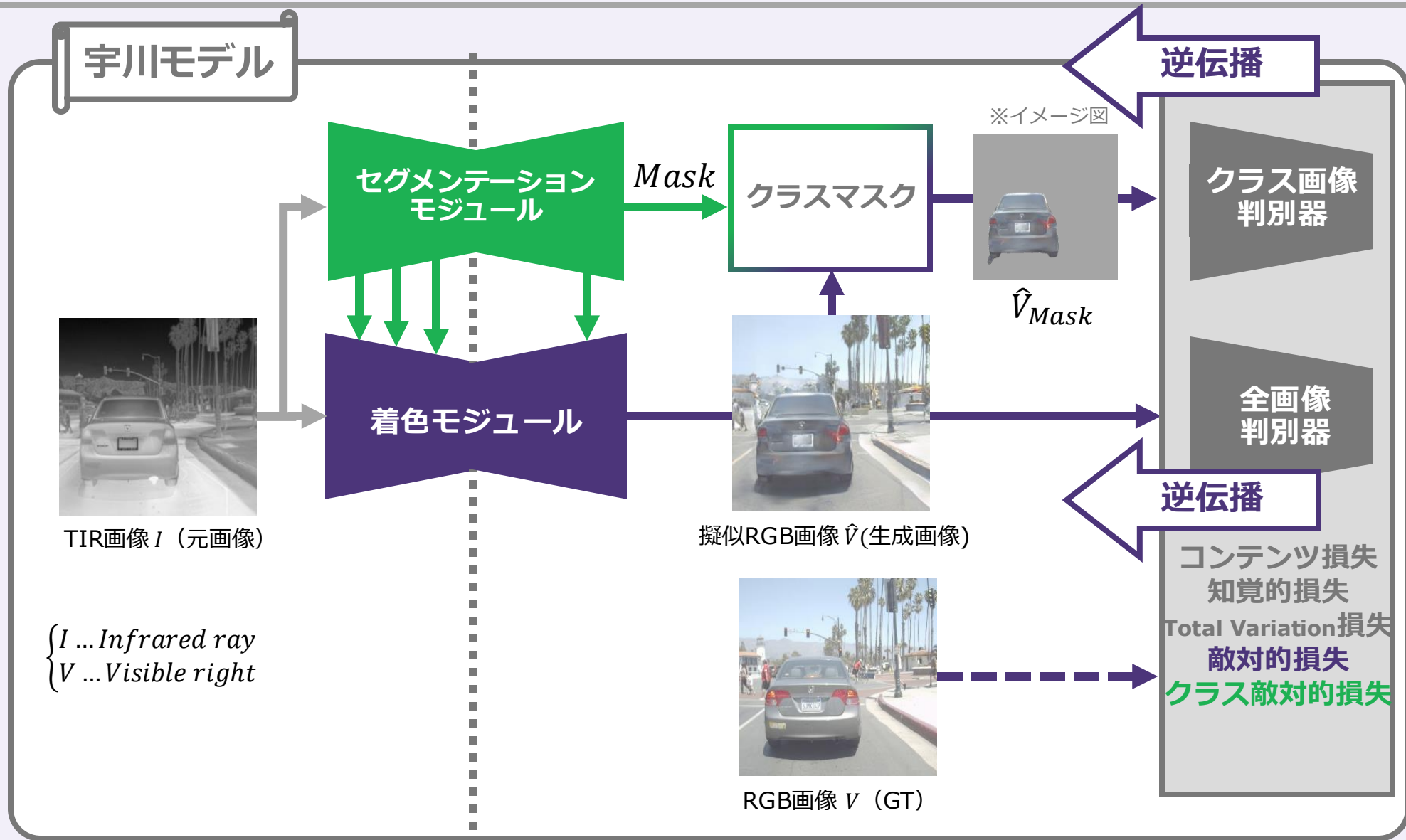


2. 関連研究 一宇川モデル

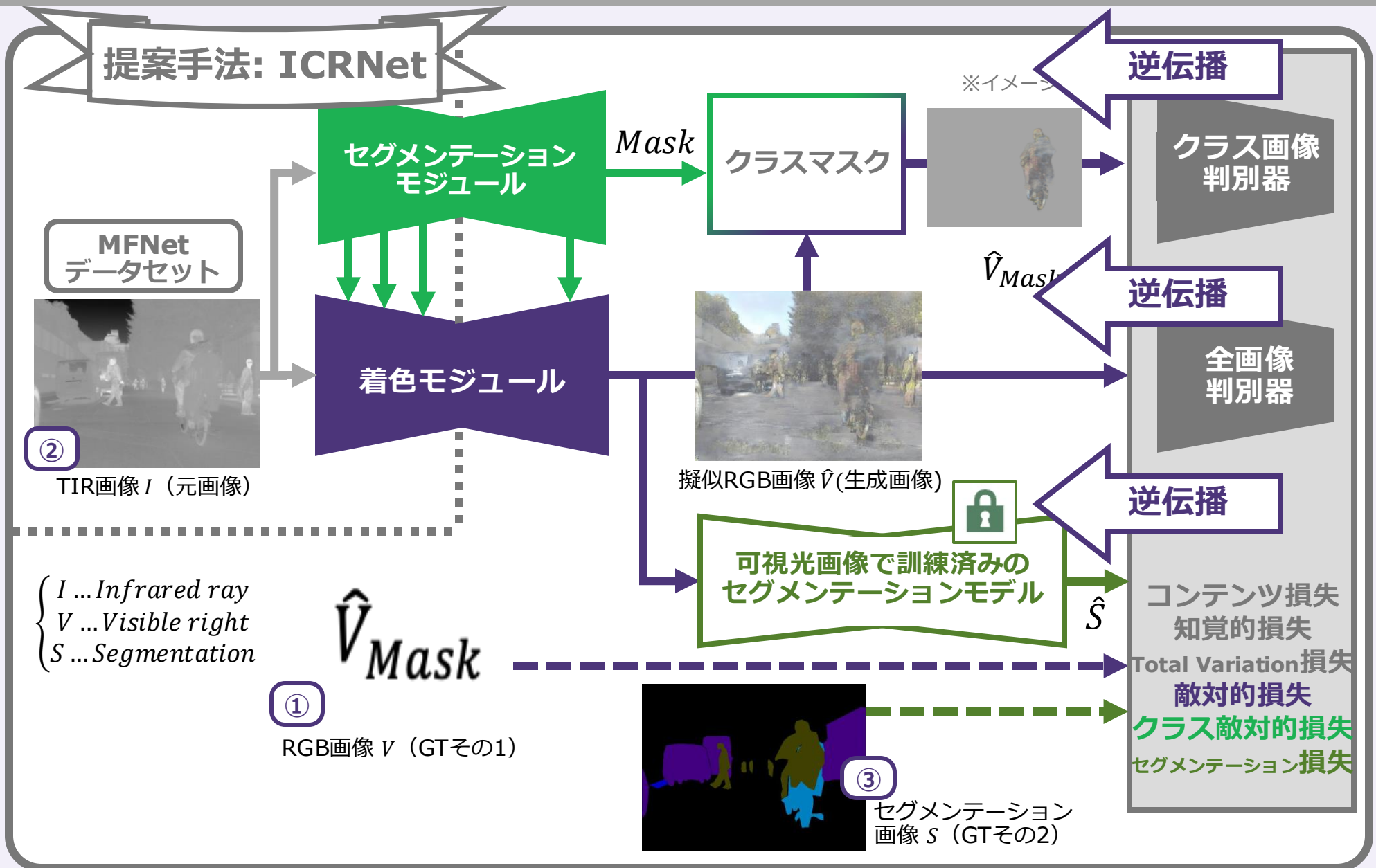
GANを利用した従来の着色モデル



2. 関連研究 一宇川モデル

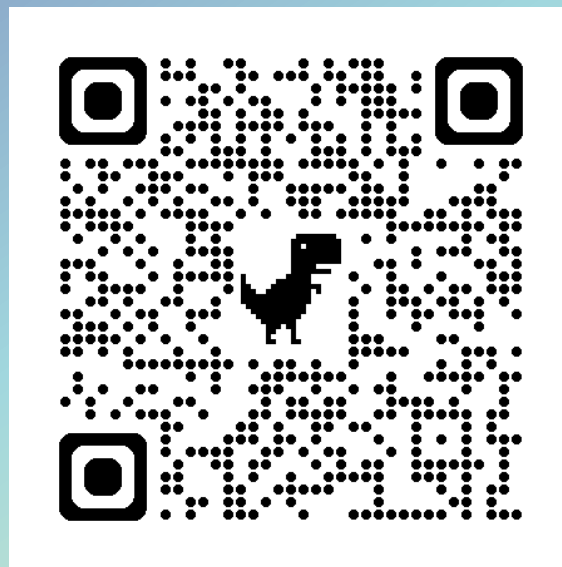


3. 提案手法



おわり

ご清聴ありがとうございました!



We eagerly await
YOUR [feedback](#)!

終
制作・著作

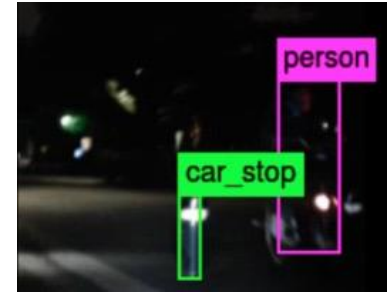


- [1] **FLIR Systems** homepage <https://www.flir.jp/>
- [2] **Satoshi Ugawa**. A Study on Thermal Infrared Image Colorization Based on Semantic Information, 2024.
- [3] **X. Kuang et al.** ‘Thermal infrared colorization via conditional generative adversarial network’, Infrared Physics & Technology, 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/
- [5] **P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros**, “Image-To-Image Translation With Conditional Adversarial Networks”.

修士での展望, 研究アイデア

学部研究を継続

- ・ セグメンテーション以外のタスクへの汎化
= MFNetに含まれている**物体検出**データの利用
- ・ 宇川モデルの軽量化



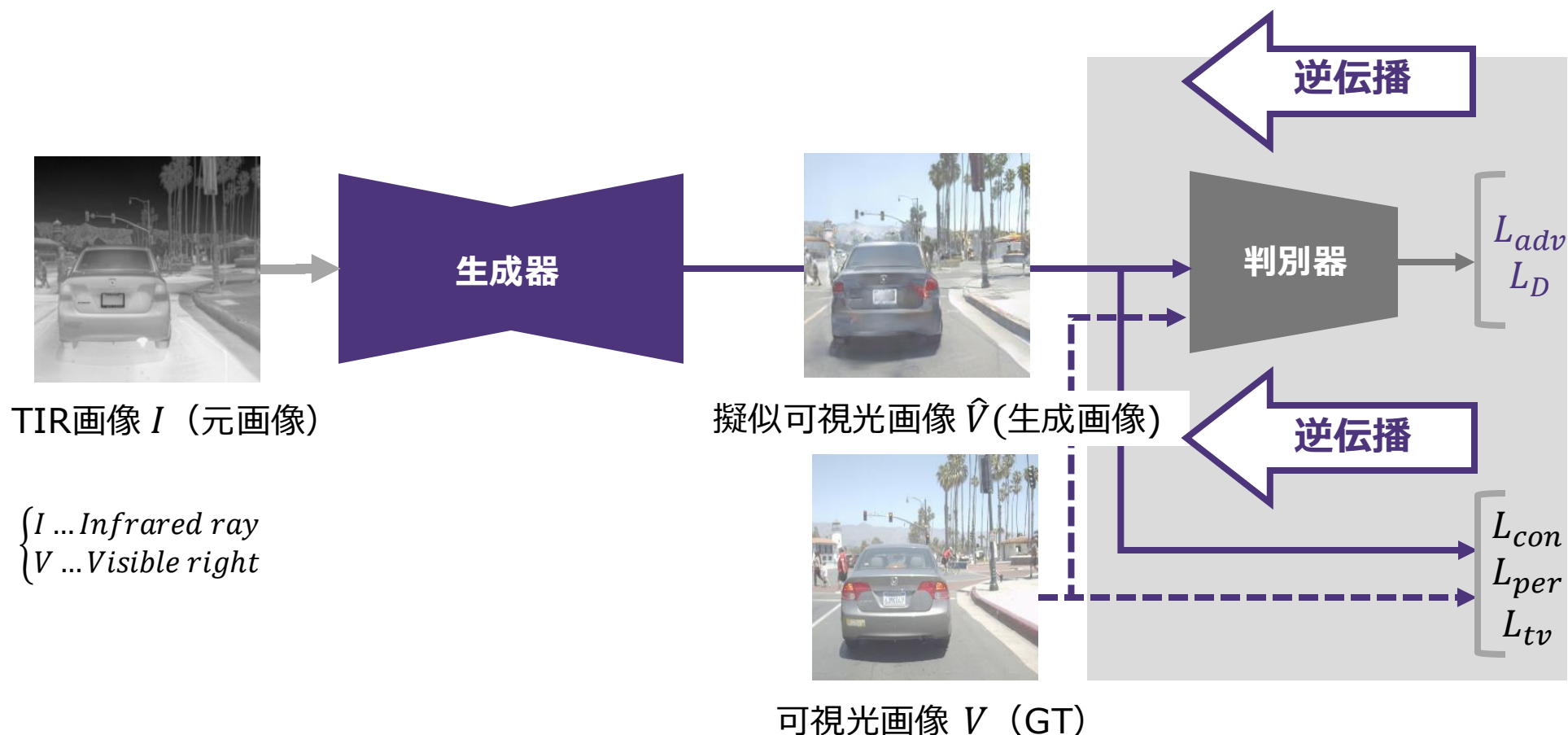
https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/det_result.png

色彩に関連のある他のテーマ

- ・ カラー画像における色が持つ情報量...モノクロに変換すると失われる意味情報の大きさを評価する
- ・ パッケージ等の表示が色弱者にとってどの程度見やすいかを定量的に評価する
- ・ 気持ちやテーマを単語で入力すると, それに従ったスライドの配色を提案するサービスの実装

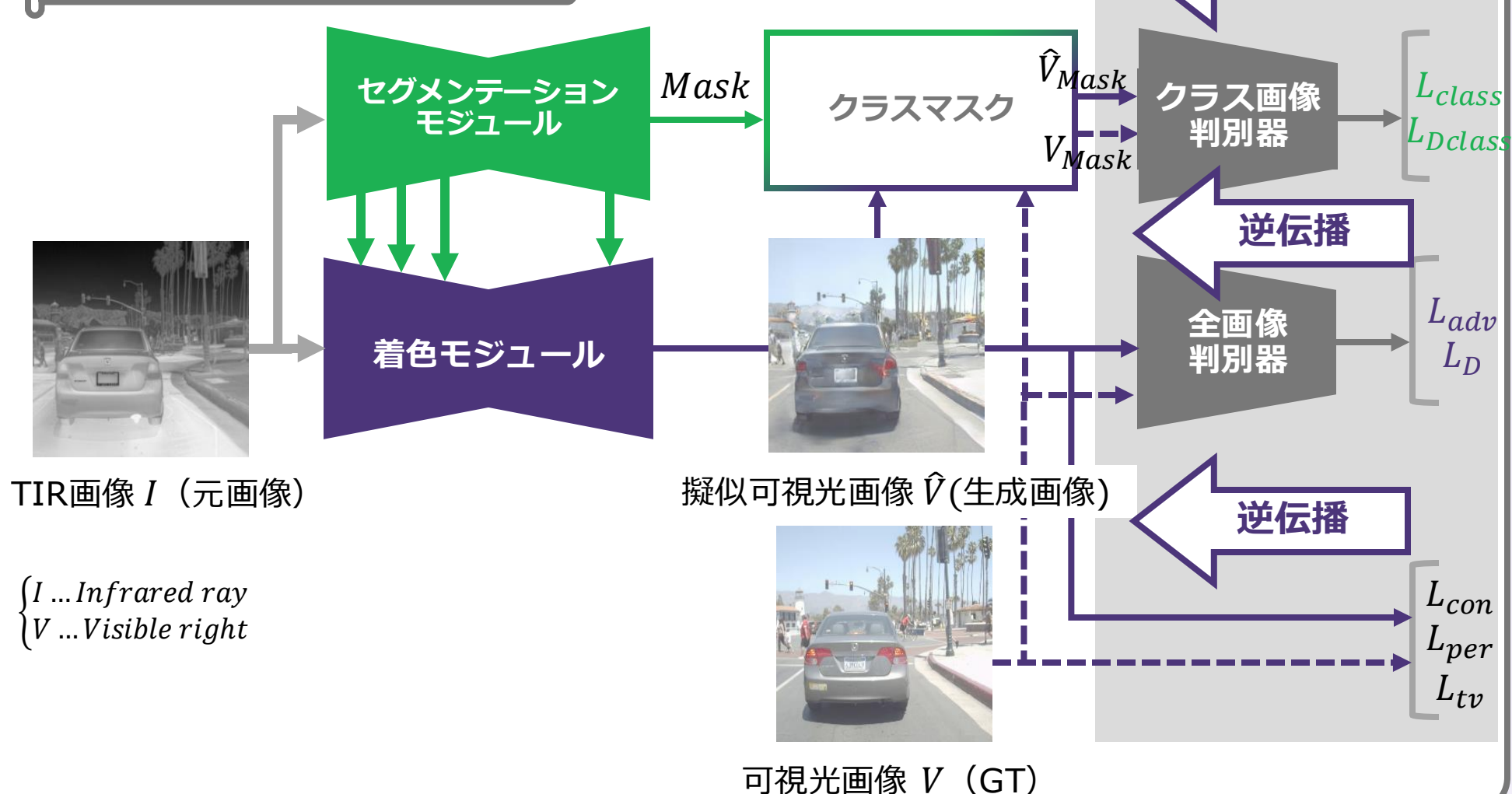
補助資料 ～損失の計算～

GANを利用した従来の着色モデル ～損失計算～

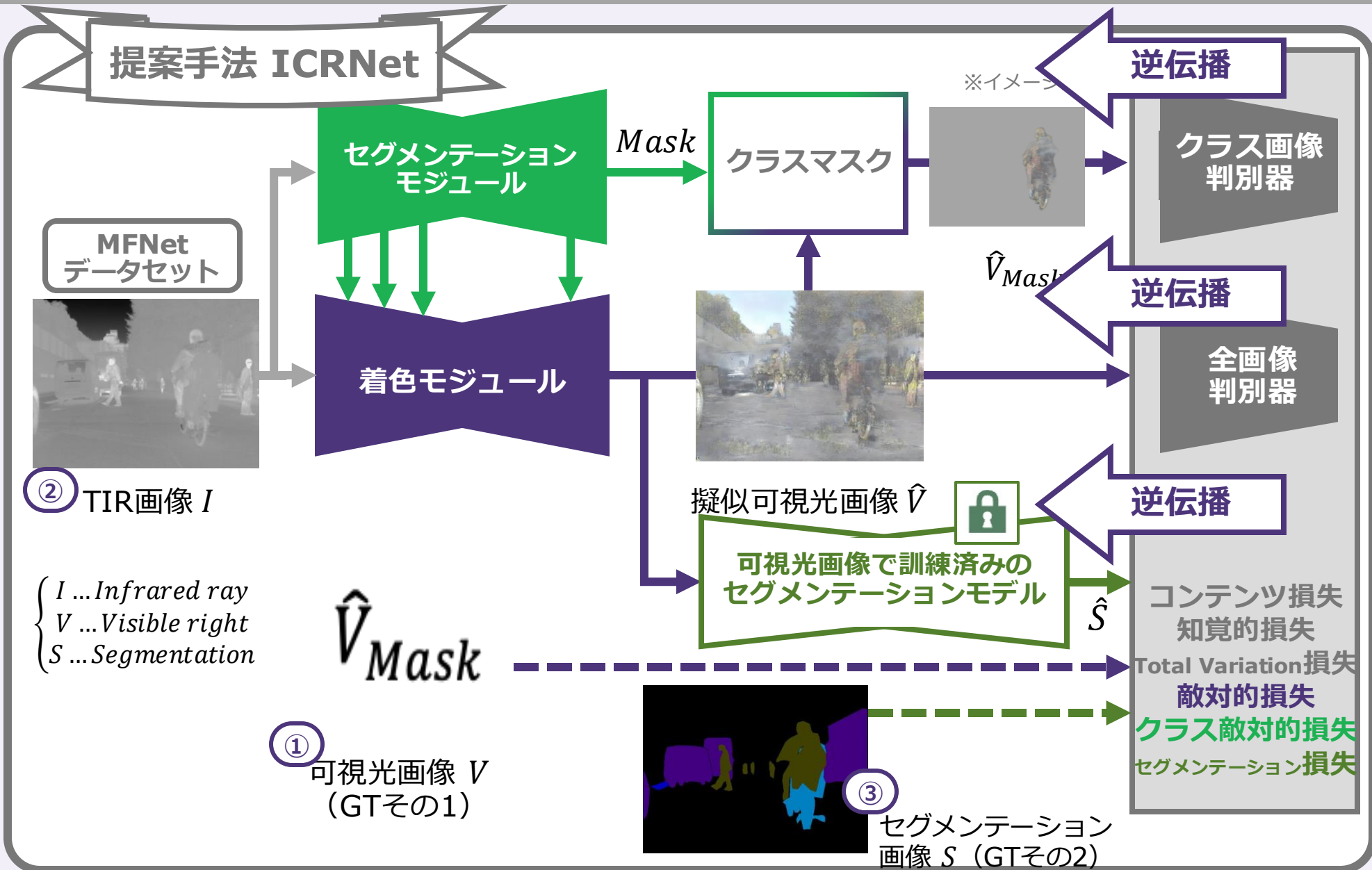


補助資料 ～損失の計算～

宇川モデル ～損失計算～

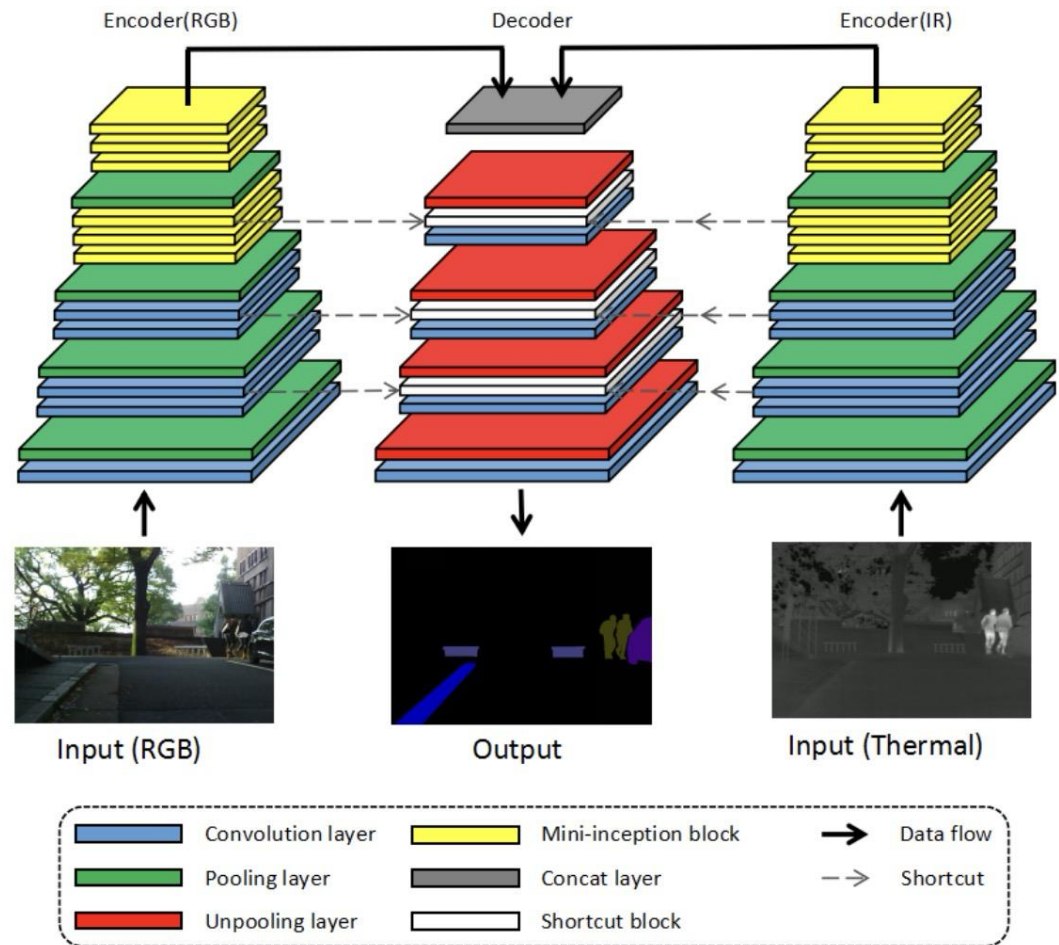


補助資料 ～損失の計算～

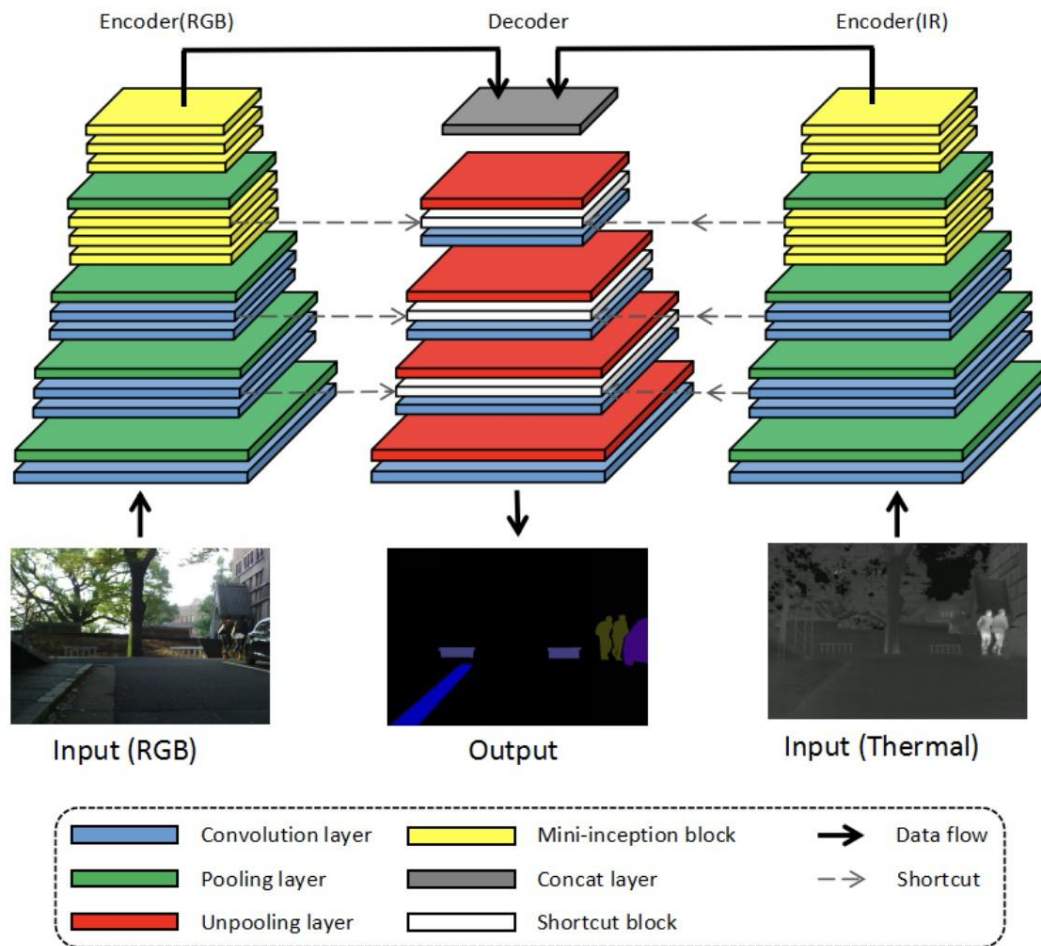


補助資料 ～MFNet～

マルチスペクトル画像は、ネットワークに渡される前にRGB画像とIR画像に分離され、その後、プロセスは各ステップを別々にエンコードする。RGBエンコーダーとIREンコーダーの出力は、デコード・ステップで融合される。エンコーダーの下位層の情報は、ショートカットを使ってデコーダーの上位層に追加される。このモデルはエンドツーエンドで学習できる。ミニインセプションとショートカットブロックの詳細を図2に示す。



補助資料 ～MFNet～



補助資料 ～宇川モデル 生成器の損失関数～

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

- 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

