視線情報を利用した一貫学習ベースによる自動運転制御の高精度化

○森 啓介†, 平川 翼†, 山下 隆義†, 藤吉 弘亘†
†: 中部大学
mkei@mprg.cs.chubu.ac.jp

概要：一貫学習方式の自動運転制御手法は，車載カメラ画像を CNN へ入力して車両制御値を直接推定する．入力画像には運転制御に関係する前方車両や歩行者などの情報だけでなく，道路外の建物などの運転制御に関係のない情報が含まれており，運転制御モデルの学習を困難にしている．そこで本研究では，自動運転制御に必要な情報を抽出するために，人の視線情報を運転制御に応用する．提案手法では，人の視線情報を直接入力するのではなく，VGG をベースとした視線推定モデルを利用して画像上の重要な領域を予測する．そして，視線推定モデルから獲得した視線情報を運転モデルの中間特徴に追加することで，自動運転制御に必要となる重要な領域を考慮した制御値推定が可能となる．CARLA シミュレータを用いた評価実験において，提案手法は，視線情報を用いない場合と比べて運転精度が向上することを確認した．
＜キーワード＞ 深層学習，自動運転，機械学習

1. はじめに

自動運転制御のアプローチは，モジュール方式と一貫学習方式に分類することができる．モジュール方式のアプローチでは，障害物検出や信号認識などの自動運転制御で必要となる処理をモジュールに分けて実行することで，周辺環境を理解し経路計画を行う．そして，車両の制御値を決定することで自動運転制御を実現している[1][2]．しかし，各モジュールの結果を統合する必要があり，詳細にタスクを分けることで自動運転システムが複雑化してしまう問題がある．一方で，画像分野における Convolutional Neural Network (CNN)の発展により，車載カメラ画像から制御値を直接推定する，一貫学習方式の自動運転制御アプローチが注目を集めている[3][4][5]．一貫学習方式の自動運転制御では，人間が実際に運転して収集した車両制御値を教師データとして模倣学習を行う．これにより，モジュール方式のようにタスクを明示的に指示することなく，運転に必要な情報を抽出して，簡素なシステムでの自動運転制御が実現可能である．しかし，入力する車載画像には運転制御に関係する前方車両や歩行者などの情報だけでなく，関係しない道路外の建物などの情報が多く含まれており，運転モデルの精度向上を困難にしている．一方で，人間は眼球運動によって必要な情報を中心窩に捉えることで効率的に視覚的な認識を行う．この視覚システムを計算モデルで再現することで，画像認識や物体検出タスクへの応用が期待されている．

そこで本研究では，視線情報を運転制御に利用する手法を提案する．提案手法では，CARLA シミュレータで作成した視線データセットにより視線推定モデルをあらかじめ用意する．そして，本モデルを用いて運転制御に必要な視線情報を予測し，その視線情報を運転モデルに利用することで自動運転の精度向上を図る．

2. 関連研究

本章では，一貫学習方式による自動運転制御および，視線推定に関する従来手法について述べる．

2.1. 一貫学習方式の自動運転制御

自動運転制御を物体検出や信号機認識などのタスクに分割して実行するモジュール方式の手法に代わるアプローチとして，車載画像から車両制御値を直接推論する一貫学習方式の自動運転手法が数多く提案されている．Pomerleau らの手法[6]では，運転画像をニューラルネットワークに入力して走行方向を推定することで，この研究の可能性を示した．Bojarski らは，簡素な構造の CNN を用いた運転制御モデル(PilotNet)を提案しており，人による運転を教師信号として高精度なステアリング制御を実現している[3]．Xu らは，LSTM を運転制御モデルに導入した FCN-LSTM を提案し，時系列を考慮しながら未来の行動ラベル(右左折, 直進等)の推論を行っている[4]．FCN-LSTM は，学習時に運転制御タスクに加えて，セマンティックセグメンテーションを補助タスクとして行うことで，コン

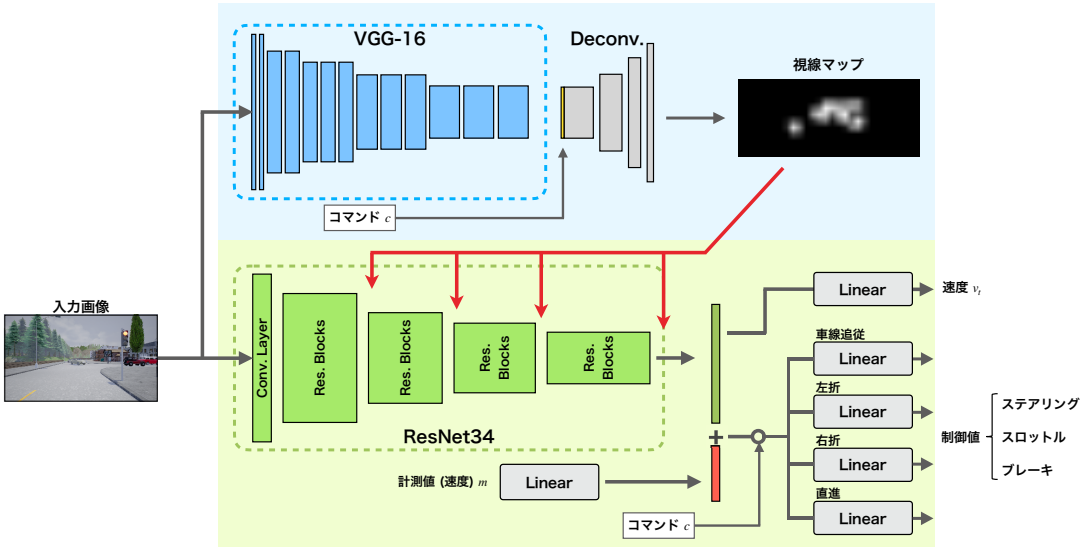

図 1：提案手法のネットワーク構造

テキスト情報を学習することが可能である．一方で，補助タスクを同時に学習するためには追加のアノテーションが必要である．そこで Hou らは，補助タスクを行う学習済みネットワークを用意し，運転制御ネットワークの中間特徴量を補助ネットワークの特徴に近似するよう学習する手法を提案している[7]．これにより追加のアノテーションをすることなく，他タスクを利用することを可能としている．また，モデルの説明可能性の問題に対して，ネットワークの判断根拠の視覚的または言語的に説明する手法も考案されている[8][9][10]．

実世界での走行実験はコストと事故の危険性があるという点から，自動運転向けシミュレータを活用した研究もされている[5][11][12]．代表的なシミュレータである CARLA[13]を用いた手法に Conditional Imitation Learning(CIL) [5]がある．CIL では，直進や右左折といった目的地までの運転指示であるコマンド情報を与えている．コマンドの数だけ出力ブランチを用意し，コマンド情報に対応したブランチを選択することで，一貫学習方式の自動運転において困難であった交差点での高精度な制御を実現している．

2.2. 視線推定

人間の視線情報には，特定のタスクにおけるその人の意図や意思決定に関する情報が含まれている．そのため，自動運転や運転支援の研究分野において，視線情報の活用が期待されており，運転シーンにおける大規模な視線データセットの作成や視線推定モデルの考案がされている．代表的なデータセットに DR(EYE)VE がある[14]．DR(EYE)VE データセットは，様々な条件下での運転時における運転手の視線を記録した 6 時間に及ぶデータセットであり，以降の研究では，このデータセットを用いた視線推定モデルが提案されている[15][16]．

視線情報を一貫学習方式の自動運転に取り入れた手法もいくつか提案されている[17][18][19]．視線マップを用いた入力画像へのマスク処理[18]や，注視箇所に基づいた多解像度画像の入力[19]など，視線推定モデルから得られた視線情報を制御値推定に用いることで，運転制御の向上を図っている．ただし，文献[18][19]の評価実験では絶対平均誤差(MAE)などによるオフライン評価のみである．文献[23]では，オフライン評価の性能が，オンラインでの走行性能に必ず寄与するとは限らないことが実験により示されている．そのため，視線情報の活用による実際の効果を検証するためには，走行評価を行う必要がある．

3. 提案手法

従来の自動運転手法では，入力画像に前方車両や歩行者などの重要な情報だけでなく，道路外の建物などの不要な情報が多く含まれる中で制御値推定を行う必要があった．本研究では，人の視線情報を利用することで，運転に必要な情報をより重視する運転制御モデルを提案する．提案するモデルのネットワークを図 1 に示す．

3.1. 視線推定

本研究では，人の視線情報を直接計測するのではなく，視線推定モデルを用いて運転タスクで重要となる注視領域を推定する．視線推定モデルのネットワークはエンコーダおよびデコーダから構

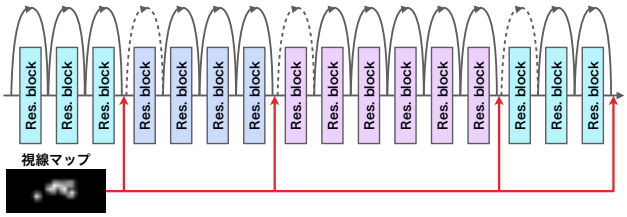

図 2：視線マップの詳細な入力箇所

成される．エンコーダには VGG-16[21]の畳み込み層を，デコーダには 4 層のデコンボリューション層を用いる．VGG-16 には ImageNet[20]による事前学習済みモデルを用いる．交差点の場面では，進行方向によって注視する箇所が大きく異なる．そのため，運転指示であるコマンド情報cをネットワークに追加することで，運転指示に合わせた視線推定を学習する．コマンド情報は，エンコーダから獲得された特徴マップと同サイズにリサイズし，チャンネル方向に結合して追加する．視線推定モデルの学習には，式(1)に示す Binary Cross Entropy (BCE) Loss を用いる．ここで，G，$\hat{G}$はそれぞれ視線マップの真値と予測値である．

$$L_g = \frac{1}{W}\frac{1}{H}\sum_{i}^{W}\sum_{j}^{H}\left(-G_{i,j}\log\hat{G}_{i,j} - (1-G_{i,j})\log(1-\hat{G}_{i,j})\right) \quad (1)$$

3.2. 運転制御モデル

運転制御モデルには，従来手法である CILRS[11]をベースラインとして用いる．CILRS では，画像を ResNet34[24]の畳み込み層を用いた特徴抽出部に入力することで特徴マップを獲得する．画像以外の情報も考慮して制御値推定を行うため，計測値である車両速度を全結合層に入力して車両速度に関する特徴量を獲得する．これらの特徴量を出力部に入力することで車両制御値を出力する．出力部では，全結合層で構成された複数ブランチの中から，運転指示（右左折，直進等）であるコマンド情報cに対応するブランチを選択する．そして，運転指示に応じた車両制御値を出力する．また，CILRS ではマルチタスクとして特徴マップから車両速度の推定を行う．

視線情報を考慮するため，式(2)に示すように特徴抽出部の中間特徴マップ$F(x)$に対して推定した視線マップGを画素毎に乗算する．そして，注視領域以外の特徴量が消失するのを防ぐため，乗算前の特徴マップを加算する．

$$F'(x) = F(x)\cdot G + F(x) \quad (2)$$

特徴抽出部の下位層から上位層では，抽出される特徴が異なる．そのため，図 2 に示すように下位層から上位層の内の 4 箇所に視線情報を追加することで，視線情報の影響をより大きくする．

運転制御モデルの学習では，損失関数としてモデルの各出力に対して L1 Loss 関数を用いる．ここでモデルの出力は，車両速度とステアリング，スロットル，ブレーキである．

4. 評価実験

提案手法の有効性を調査するため，視線情報の有無による運転精度の比較を行う．

4.1. データセット

本実験では，自動運転シミュレータである CARLA にて作成された CARLA100 データセット[11]を運転制御モデルの学習に用いる．CARLA100 データセットは，車両の位置や信号機の状態などのシミュレータ内部の情報からルールベースで自動運転を行い，100 時間におよぶ運転データを自動的に収集したデータセットである．また，模倣学習における問題点である共変量シフトに対する頑健性を高めるため，別視点カメラの追加[3]および noise injection[5][22]によるデータ拡張が施されている．

本研究では，視線推定モデルを学習させるため，新規に視線データセットを作成する．視線データセットは CARLA100 データのサブセットとなっており，各天候，運転指示に対応したシーンの中からランダムで選択した 2,000 フレームのデータに対して視線情報を付与する．1 フレーム毎に十分な視線情報が含まれるようにするため，静止画に

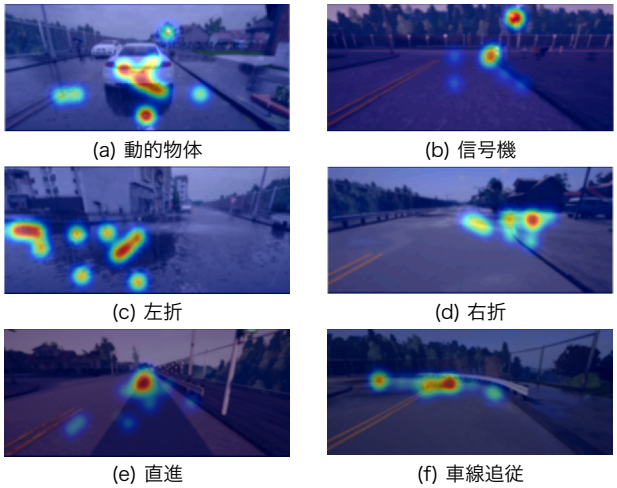

図 3：記録した視線データの例

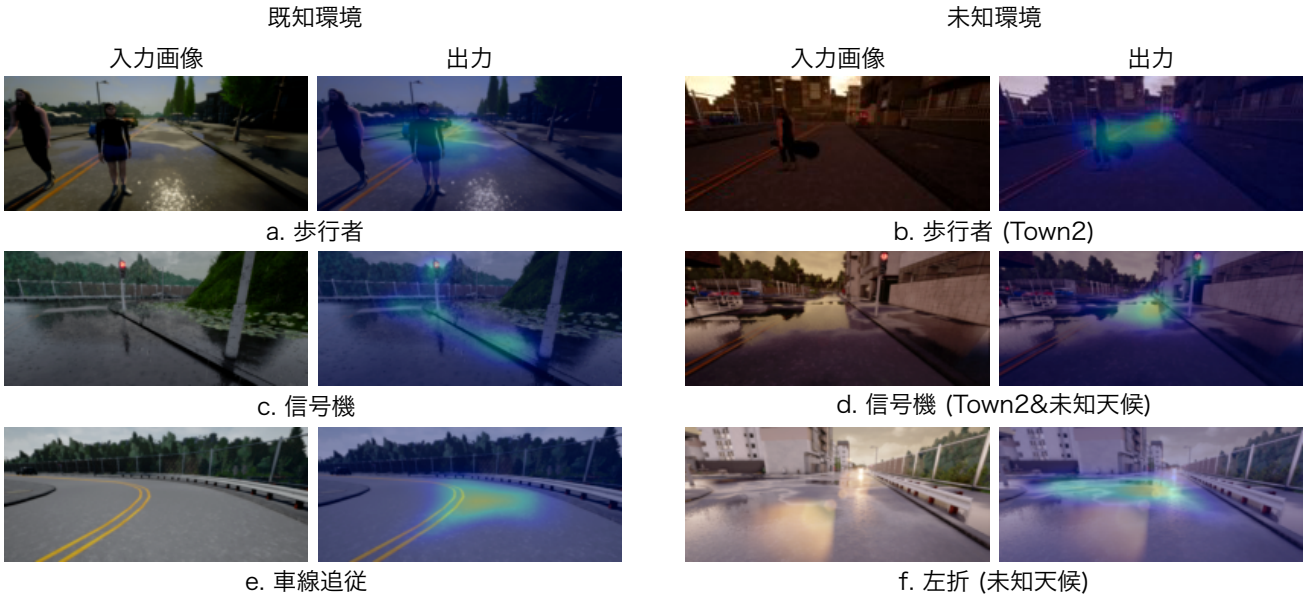

図 4：視線推定の結果

対する視線情報を記録する．視線情報の取得には Tobii Pro X3 130 を使用し，1 枚の静止画をディスプレイに 5 秒間表示することでデータを収集する．運転に必要な情報を正確に抽出するため，被験者には，表示された運転シーンに対して必要な運転操作を考慮するよう指示を与えて視線を記録する．作成したデータセットの例を図 3 に示す．

4.2. 実験概要

視線推定モデルは， 2,000 枚の視線データを用いて学習を行う．入力は400 × 176ピクセルにリサイズした RGB 画像を用いる．最適化手法には MomentumSGD を用い，学習率を 0.03 とする．運転制御モデルは，CARLA100 データセットの内 10 時間のサブセットを学習データとして，200 × 88 ピクセルの RGB 画像を用いる．最適化手法には Adam を使用し，学習率を 0.0002 とする．学習は 2 段階に分けて行う．はじめに視線推定モデルを学習した後，視線推定モデルの重みを固定した状態で運転制御モデルの学習を行う．

評価実験では，CARLA 上での走行実験により手法の有効性を検証する．実験環境には CARLA 0.8.4 を使用し，NoCrash Benchmark[11]により実験評価を行う．このベンチマークでは，異なる環境および難易度において目的地までの自律走行することで運転性能を評価する．走行するマップは Town01 と Town02 で，Town02 は学習データに含まれていないマップとなっている．また，天候条件についても学習データに含まれる条件(Clear noon, Clear noon after rain, Heavy raining noon)および未知の条件(Cloudy noon after rain, Soft raining sunset)が用意されている．交通状況による難易度は，自動車や歩行者といった動的物体の数により Empty town, Regular traffic, Dense traffic の 3 種類に分けられている．これらの条件により，異なる環境や交通状況における運転性能を評価することが可能である．他物体への衝突または制限時間を超えた場合，エピソードが失敗となり，目的地へ到達することでエピソード成功となる．

4.3. 視線推定の定性的評価

はじめに，視線推定モデルの定性的な評価を行う．視線推定結果の例を図 4 に示す．ここで，左が学習に含まれる環境，右が未知の環境での推定結果である．上段の(a), (b)のような歩行者が道路を横断しているシーンにおいて，(a)の既知環境では衝突の危険がある右側の歩行者に対して視線マップが反応していることがわかる．(b)の Town02 においても，進行方向である消失点だけでなく，横断中の歩行者に対しても期待通りの視線マップが予測されていることがわかる．中段の信号機に接近しているシーンにおいては，(c)既知環境と(d)未知環境の両方で，スロットルの制御に関わる信号機を注視していることがわかる．下段(e)の車線追従シーンでは，走行レーンの進行方向であるカーブの行き先を注視している．(f)の交差点のシーンでは，左折の運転指示が与えられており，視線推定結果は左折する際の進行方向を注視している．(e), (f)から運転指示に応じて先の進行方向に対する視線が推定されていることが確認できた．これらの結果より，多様な走行環境において運転制御

表 1 NoCrash ベンチマークにおけるエピソード成功率の比較結果[%]

	Training conditions		New weather		New town		New town & weather	
	CILRS	Ours	CILRS	Ours	CILRS	Ours	CILRS	Ours
Empty	92	**95**	92	**96**	54	**66**	**72**	66
Regular	63	**77**	**64**	60	29	**46**	**44**	42
Dense	15	**25**	8	**28**	8	**12**	8	**12**

表 2 動的物体との事故率の比較結果[%]

		Training conditions		New weather		New town		New town & weather	
		CILRS	Ours	CILRS	Ours	CILRS	Ours	CILRS	Ours
Empty	Col. Pedestrian	0	0	0	0	0	0	0	0
	Col. Vehicle	0	0	0	0	0	0	0	0
Regular	Col. Pedestrian	9	**4**	**6**	14	**9**	**9**	**6**	10
	Col. Vehicle	21	**11**	22	**18**	30	**16**	24	**16**
Dense	Col. Pedestrian	21	**16**	**14**	16	**10**	16	**14**	20
	Col. Vehicle	57	**47**	68	**48**	68	**58**	70	**52**

の判断に必要と考えられる領域を注視できていることが確認できた.

4.4. 走行実験結果

次に，運転制御モデルの評価を行う．表 1 に NoCrash Benchmark による走行実験のエピソード成功率を示す．表 1 より，ほとんどの交通状況および環境条件において従来手法より成功率が向上しており，特に Training condition および New town では全ての交通状況において成功率が改善されたことが確認できた．また，Training condition の regular traffic と dense traffic，New town の regular traffic の成功率は，それぞれ 14 ポイント，10 ポイント，17 ポイントと大きく改善されていることから，提案手法が従来手法と比べて動的物体に対して適切な運転制御ができていることが考えられる．しかし，New weather と New town & new weather において，成功率が低下していることがわかる．どちらの条件にも未知の天候が含まれていることから，視線推定モデルは未知の天候に対する頑健性が不足していることが原因として考えられる．同一の実験における事故率を表 2 に示す．ここで，Col. Pedestrian および Col. Vehicle はそれぞれ歩行者との衝突による事故と車との衝突による事故を表す．従来手法と比べて，自動車との事故率はどの実験条件においても大幅に改善されていることがわかる．これにより，視線情報を付与することで自動車に対する反応が改善され，動的物体が存在する運転シーンにおける精度向上に貢献していることが確認できた．一方で，歩行者との事故率は未知の条件が含まれる環境においては，改善が確認されなかった．これは，学習データセット内の自動車が含まれるシーンと歩行者が含まれるシーンの発生頻度が異なり，視線推定モデルが自動車をより認識できていたためであると考えられる.

5. おわりに

本研究では，視線情報を活用した一貫学習方式の自動運転手法を提案した．視線推定モデルから獲得した視線情報を運転制御モデルの中間特徴に追加することで，運転シーンにおける重要な領域を強調し，シミュレータ上での走行実験において運転精度が改善できることを確認した．今後の展望として，歩行者シーンの運転性能向上や，視線推定モデルの特徴量を運転制御モデルに利用した手法の検討が挙げられる.

参考文献

[1] C. Urmson, J. Anhalt, D. Bagnell, C. Baker, R. Bittner, M. Clark, J. Dolan, D. Duggins, T. Galatali, C. Geyer, et al., “Autonomous driving in urban environments: Boss and the urban challenge,” Journal of Field Robotics, vol.25, no.8, pp.425-466, 2008.

[2] E. Guizzo, “How google’s self-driving car works,” IEEE Spectrum Online, vol.18, no.7, pp.1132-1141, 2011.

[3] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B,

Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, K. Zieba, “End to End Learning for Self-Driving Cars,” arXiv preprint arXiv:1704.07911, 2016.

[4] H. Xu, Y. Gao, F. Yu, T. Darrell, “End-to-end learning of driving models from large-scale video datasets,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2174–2182, 2017.

[5] F. Codevilla, M. Müller, A. López, V. Koltun, “End-to-end Driving via Conditional Imitation Learning,” International Conference on Robotics and Automation (ICRA), pp.4693-4700, 2018.

[6] D. Pomerleau, “ALVINN: An autonomous Land Vehicle in a Neural Network,” Neural Information Processing Systems (NeurIPS), pp.305–313, 1989.

[7] Y. Hou, Z. Ma, C. Liu, C. C. Loy, “Learning to Steer by Mimicking Features from Heterogeneous Auxiliary Networks,” Association for the Advancement of Artificial Intelligence (AAAI), pp.8433-8440, 2019.

[8] M. Bojarski, P. Yeres, A. Choromanska, K. Choromanski, B. Firner, L. D. Jackel, U. Muller, “Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car,” arXiv preprint, arXiv:1704.07911, 2017.

[9] J. Kim, J. Canny, “Interpretable Learning for Self-Driving Cars by Visualizing Causal Attention,” nternational Conference on Computer Vision (ICCV), pp.2942-2950, 2017.

[10] J. Kim, A. Rohrbach, T. Darrel, J. Canny, Z. Akata, “Textual Explanations for Self-Driving Vehicles,” European Conference on Computer Vision (ECCV), pp.577-593, 2018.

[11] F. Codevilla, Eder Santana, A. M. López, A. Gaidon, “Exploring the Limitations of Behavior Cloning for Autonomous Driving,” International Conference on Computer Vision (ICCV), pp.9329-9338, 2019.

[12] D. Chen, B. Zhou, V. Koltun, P. Krähenbühl, “Learning by Cheating,” Conference on Robot Learning (CoRL), pp.66-75, 2020.

[13] A. Dosoviskiy, G. Ros, F. Codevilla, A. López, V. Koltun, “CARLA: An Open Urban Driving Simulator,” Conference on Robot Learning (CoRL), pp.1-16, 2017.

[14] S. Alletto, A. Palazzi, F. Solera, S. Calderara, R. Cucchiara, “DR(eye)VE: A Dataset for Attention-Based Tasks with Applications to Autonomous and Assisted Driving,” IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.54-60, 2016.

[15] A. Palazzi, F. Solera, S. Calderara, S. Alletto, R. Cucchiara, “Learning where to attend like a human driver,” Intelligent Vehicles Symposium (IV), pp. 920–925, 2017.

[16] A. Tawari and B. Kang, “A computational framework for driver’s visual attention using a fully convolutional architecture,” Intelligent Vehicles Symposium (IV), pp. 887–894, 2017.

[17] Y. Chen, C. Liu, L. Tai, M. Liu, B. E. Shi, “Gaze Training by Modulated Dropout Improves Imitation Learning,” International Conference on Intelligent Robots and Systems (IROS), pp.7756-7761, 2019.

[18] A. Makrigiorgos, A. Shafti, A. Harston, J. Gerard, A. A. Faisal, “Human Visual Attention Prediction Boosts Learning & Performance of Autonomous Driving Agents,” arXiv preprint arXiv: 1909.05003, 2019.

[19] Y. Xia, J. Kim, J. Canny, K. Zipser, T. C. Bajo, D. Whitney, “Periphery-Fovea Multi-Resolution Driving Model Guided by Human Attention,” Winter Conference on Applications of Computer Vision (WACV), pp.1767-1775, 2020.

[20] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L. F. Fei, “ImageNet: A large-scale hierarchical image database,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248-255, 2009.

[21] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556 (2014).

[22] M. Laskey, J. Lee, R. Fox, A. Dragan, K. Goldberg, “DART: Noise Injection for Robust Imitation Learning,” Conference on Robot Learning (CoRL), 2017.

[23] F. Codevilla, A. M. Lopez, V. Koltun, A. Dosovitskiy, “On offline evaluation of vision-based driving models,” European Conference on Computer Vision (ECCV), pp.236–251, 2018.

[24] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778, 2016.

森啓介：2019 年 中部大学工学部情報工学科卒業．現在中部大学大学院修士課程在学中，一貫学習による自動運転制御の研究に従事．

平川翼： 2013 年 広島大学大学院博士課程前期修了，2014 年 広島大学大学院博士課程後期入学，2017 年 中部大学研究員 (~2019)，2017 年 広島大学大学院博士後期課程修了．2019 年 中部大学特任助教．コンピュータビジョン，パターン認識， 医用画像処理の研究に従事．

山下隆義：2002 年 奈良先端科学技術大学大学院大学博士前期課程修了．2002 年 オムロン株式会社入社，2009 年 中部大学大学院博士後期課程修了(社会人ドクター)，2014 年 中部大学講師，2017 年より同大学准教授．人の理解に向けた動画像処理，パターン認識・機械学習の研究に従事．

藤吉弘亘：1997 年 中部大学大学院博士後期課程修了．1997～2000 年 米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow．2000 年 中部大学講師．2004 年より同大学教授．2005～2006 年 米カーネギーメロン大学ロボット工学研究所客員研究員，計算機視覚，動画像処理，パターン認識・理解の研究に従事．