# Next-ViTとBiFPNによる車載カメラ映像からのマルチタスクの高精度化

○張 陳雨 †, 平川 翼 †, 山下 隆義 †, 藤吉 弘亘 † †:中部大学

zhang1121@mprg.cs.chubu.ac.jp

概要:自動運転システムには多種多様なタスクが存在するため、マルチタスク学習は非常に効果的である. マルチタスク学習におけるバックボーンでは下流タスクに共通した特徴を抽出する.しかし、下流タスク により必要な特徴レベルが異なるため、各下流タスクに合わせた特徴抽出が必要である.そこで本研究で は、車載カメラ映像を用いた自動運転タスクにおいて CNN と Transformer を活用した Next-ViT と、特 徴融合手法 BiFPN を導入したマルチタスクモデルを提案する.自動運転タスクによる評価実験から、提 案手法を用いることでマルチタスク学習における精度向上を確認した.

<キーワード>マルチタスク学習,自動運転支援,Next-ViT,BiFPN

# 1. はじめに

自動運転技術は認知,判断,操作をドライバー の代わりにシステムにより自動化する技術である. この技術開発には,交通事故の削減,過疎地域で の新たな移動手段の確立など社会課題を解決する とともに,モビリティ産業に大きなインパクトを 与えることが期待されている.

自動運転システムは道路上の車両や歩行者の認 識,交通信号や標識の理解,適切な車両制御など, 複数のタスクを高精度にこなす必要がある.複数 タスクを同時に学習する手法としてマルチタスク 学習がある.マルチタスク学習は,互いに関連し た複数のタスクを単一モデルにより同時に学習す ることで精度向上を図る手法である [1, 2, 3].自 動運転タスクで解くべき物体検出やセグメンテー ションなどは,互いに密接に関連しているためマ ルチタスク学習が効果的である [4, 5].

しかし,従来のマルチタスク学習における学習 プロセスには問題点がある.マルチタスク学習で は,バックボーンネットワークにより下流タスク 間で共通した特徴を抽出する.したがって,この バックボーンネットワークにおける特徴抽出能力 が不十分である場合,全下流タスクにおける精度 低下を引き起こす可能性がある.また,下流タス クごとに必要な特徴レベルが異なるため,精度向 上には下流タスクごとに合わせた特徴が求められ る. これは自動運転タスクにおいても同様である.

そこで、本研究ではバックボーンネットワーク に CNN と Transformer を活用した Next-ViT を 採用し、下流タスクへの特徴伝播に特徴融合手法 BiFPN を導入したマルチタスクモデルを提案する. 我々の研究では自動運転タスクを対象とし、車載 カメラ画像を用いた下流タスクの高精度化を目的 とする. BDD100K データセットでの評価実験に より、自動運転におけるマルチタスクモデルの性 能向上が確認した.

#### 2. 関連研究

## 2.1. 自動運転タスク

自動運転には、車載カメラ画像から道路上の物 体検出タスク(車両,歩行者など)や運転可能領域 のセグメンテーションタスク,車道線検出タスク など複数のタスクがある.

物体検出タスクでは、1ステージ検出手法と2ス テージ検出手法がある. 1ステージ検出手法では 物体のクラススコアとバウンディングボックスを 同時に出力する. 代表的な手法としては YOLOv4 [6] が挙げられる. 2ステージ検出手法では物体候 補領域を抽出後に識別を行い、物体のクラススコ アとバウンディングボックスを出力する. 代表的 な手法としては Faster R-CNN [7] が挙げられる.

セグメンテーションタスクでは,入力画像内の ピクセルごとでクラス分類を解くため,多様なレ



図1 提案手法のネットワーク構造:BiFPN 内の P は各レベルの特徴を表している.

ベルの特徴を効率的に融合することが求められる. 代表的な手法である PSP-Net はピラミッドプーリ ングモジュールを導入し,低レベルと高レベルの特 徴を融合することで高精度な結果を獲得している [8]. ピラミッドプーリングモジュールは,Feature Pyramid Network (FPN) 内に最大プーリング を導入したモジュールである.これにより局所的 情報だけでなく大域的情報を獲得している.

車道線検出タスクは,セグメンテーションにお けるサブタスクであり,車載カメラ画像内に含ま れる車道線領域をセグメンテーションするタスク である.代表的な手法である ENet-SAD は知識蒸 留を導入することで,高精度でリアルタイムな車 道線検出を実現している [15].

# 2.2. 自動運転におけるマルチタスク学習

マルチタスク学習におけるモデル構造は Encoder-Decoder 型が主流である. Encoder-Decoder モデルでは, Encoder でバックボーン ネットワークによる下流タスク間で共通特徴を 抽出し, Decoder で Encoder により抽出した特 徴を用いて下流タスクに対する出力値を算出す る. 自動運転におけるマルチタスク学習として, Teichmann らは物体検出とセグメンテーションを 単一モデルで学習する MultiNet を提案している [4].

マルチタスクモデルは Encoder-Decoder 構造を 採用しているため, 2つの問題点がある. 1つ目は, Encoder でのバックボーンネットワークによる特 徴抽出性能が下流タスクの精度に大きく影響する 点である. バックボーンネットワークの特徴抽出 が不十分であれば、全下流タスクにおいて精度が 低下してしまう. 2つ目は, Decoder での下流タス クは必要な特徴レベルが異なる点である. Decoder では異なるタスクを解くため、Decoderへの入力と して下流タスクに合わせたレベルの特徴が必要で ある. これらの問題に対し解決を試みた手法とし て DLT-Net がある [5]. この手法は, 自動運転タ スクである物体検出とセグメンテーション, 車道 線検出を対象とし, Encoder モデルに CNN ベー スの VGG16 と FPN 構造を導入することで、 各 タスクの精度向上を報告している. つまり、 バッ クボーンネットワークに特徴抽出性能を高いモデ ルを採用し、様々な特徴レベルを考慮して特徴抽 出することができれば精度向上を実現できる.し かし、自動運転のような高難易度な下流タスクで ある場合, VGG16 では特徴抽出性能が低く, 自動 運転タスクにおいて十分な精度を見込めない.

一方で,我々は局所および大域的特徴を獲得可 能な Next-ViT [9] をバックボーンネットワークと して採用し,抽出した特徴をうまく下流タスクに適 用する特徴融合手法 BiFPN [10] を Encoder に導 入した.これにより,自動運転におけるマルチタ スクモデルでの下流タスクの精度向上を実現した.

## 3. 提案手法

本研究では,自動運転におけるマルチタスクモデ ルの Encoder に着目し, Next-ViT [9] と BiFPN [10] を導入したマルチタスクモデルを提案する.提 案手法のモデル構造を図 1 に示す.

# 3.1. Encoder

本手法における Encoder のバックボーンネット ワークには Next-ViT を採用する.またマルチタ スク学習における Decoder への入力には,各下流 タスクに合わせた特徴レベルの特徴が必要である. そこで BiFPN を導入し Next-ViT から抽出した 特徴に対し下流タスクごとに合わせた特徴融合を 行う.

Next-ViT. コンピュータービジョン領域で は Vision Transformer (ViT) [11] を代表とした Transformer を用いたモデルが高い性能を達成し ている. ViT は大域的な特徴を捉えるのに適して いるが、局所的な特徴は CNN ベースモデルほどの 特徴抽出性能は高くなく推論速度も遅い、そこで CNN による局所的な特徴抽出と Transformer によ る大域的な特徴抽出を組み合わせたモデルである Next-ViT が提案されている. Next-ViT は CNN ブロックである Next Convolution Block (NCB) と Transformer ブロックである Next Transformer Block (NTB) を用いたネットワーク構造である. 図2に Next-ViT のネットワーク構造を示す. NCB では、グループ間アテンションモジュール Group Convolution を用いた Multi-head Convolution Attention (MHCA) と Multi-Layer Perceptron (MLP) を採用し、かつスキップ接続を導 入することで強力な局所的特徴抽出能力を獲得し ている. NTB では, Efficient Multi-head Self-Attention (E-MHSA) を導入することで、 ViT のような強力な大域的特徴獲得能力を獲得してい る. NTB 内部では E-MHSA 以降に NCB と同様 の構造を導入することで,獲得した大域的特徴を 局所的特徴と融合している. このモデルは高い特 徴抽出性能を保有しているため, バックボーンネッ トーワークとして最適である.

BiFPN. 特徴融合とは, ニューラルネットワー クによる特徴抽出において, 豊富なセマンティック 情報を持つ低レベルな特徴と, 貧弱なセマンティッ



図 2 Next-ViT のネットワーク構造

ク情報を持つ高レベルな特徴を, 接続および融合 し正確な位置情報とセマンティック情報を獲得す ることである. 従来の特徴融合では, 単方向結合 やスキップ接続など, 単純な特徴マップを重ね合 わせや加算である. しかし, これでは効率的に特 徴融合を行えていない. そこで, 層ごとの特徴に 重み付けし双方向で特徴融合する BiFPN が提案 されている.

図1に提案手法における BiFPN を示す. 青矢 印は, Deconvolution layer による高レベル特徴を 低レベル特徴へアップサンプリングする処理を示 している. 茶矢印は, Convolution layer による低 レベル特徴を高レベル特徴へダウンサンプリング する処理を示している. またピンク矢印はスキッ プ接続である. 各層では同サイズに変換した様々 なレベルの特徴を重み付け和により特徴融合する. このような特徴融合操作を実施することで,より 豊富な特徴情報が獲得できる.

## 3.2. Decoder

本研究では下流タスクとして,自動運転におけ る物体検出と運転可能領域セマンティックセグメ ンテーション,車道線検出を対象とする.そのた め Decoder には検出ヘッドとセグメンテーション ヘッドを構築する.検出ヘッドでは,BiFPN から 取得した5層の特徴 P<sub>2</sub> ~ P<sub>6</sub> を用い, バウンディ ングボックスのオフセット, クラス確率, および信 頼度を予測する. セグメンテーションヘッドでは, 運転可能領域, 車道線, および背景の3クラスによ りセグメンテーションする. 低レベルな特徴は豊 富なセマンティック情報を持つため, 入力値には BiFPN から取得した5層の特徴 P<sub>2</sub> ~ P<sub>6</sub> と, 低レ ベルな特徴 P<sub>1</sub> を用いる. また, 各特徴マップのサ イズは P<sub>1</sub> と同サイズにアップサンプリングする.

#### 3.3. 損失関数

マルチタスク学習は,タスク間における学習難 易度の相違により,全タスクを効率的に学習できな い問題がある.そこで,我々は損失関数 Focal loss [12] と Tversky loss [13] を採用する.提案手法に おける損失関数 *L<sub>T</sub>otal* を式 (1) に示す.

$$L_{Total} = \alpha L_{det} + \beta L_{seg} \tag{1}$$

ここで $\alpha, \beta$ は損失間のバランスをとるためのパ ラメータである.  $L_{det}$ は物体検出タスクの損失,  $L_{seg}$ はセグメンテーションタスクの損失である. タタスクの損失 L. たず(2) に示す

各タスクの損失 *L*<sub>det</sub>, *L*<sub>seg</sub> を式 (2) に示す.

$$L_{det} = \alpha_{cls}L_{cls} + \alpha_{obj}L_{obj} + \alpha_{bbox}L_{bbox}$$

$$L_{seg} = L_{Tversky} + \gamma L_{Focal}$$
(2)

ここで  $\alpha_{cla,obj,bbox}, \gamma$  は損失間のバランスをとるた めのパラメータである.  $L_{cls} \ge L_{obj}$ は、オブジェ クトのクラスと信頼度に関する損失であり、 Focal loss に基づき設計する. Focal loss はクラス不均衡 を低減し, 誤分類しやすいサンプルに比重を置く 損失である. L<sub>bbox</sub> は,予測ボックスと真値ボック ス間の絶対値をとる smooth L1 損失により算出す る. L<sub>seg</sub>は、運転可能領域、車道線、背景のセグ メンテーションに関するマルチクラス損失であり, Focal loss  $L_{Focal}$  と Tversky loss  $L_{Tversky}$  に基づ き設計する. Tversky loss はクラス不均衡を低減 し、小領域のクラスに対しても精度を向上させる ことができる損失である.つまり, L<sub>Tversky</sub> はク ラス不均衡な問題で優れた性能を発揮し、 LFocal は、ピクセル間の分類誤差を最小限に抑えハード ラベルに焦点を当てる.

## 4. 評価実験

本実験では Berkeley Deep Drive データセット (BDD100K) [14] を用いて提案手法を評価する. 評 価方法は以下の通りである.

- 下流タスクにおける性能比較
- NextViT の有効性評価
- BiFPN の有効性評価

学習用画像 70,000 枚,評価用画像 10,000 枚とした.また,入力画像サイズは 640 × 384 にリサイズ した.

提案手法の学習フローは以下の通りである.

- Step1 セグメンテーションヘッドを凍結し, バッ クボーンと検出ヘッドを 200epoch 学習.
- Step2 バックボーンと検出ヘッドを凍結し, セグ メンテーションヘッドを 50epoch 学習.

Step3 モデル全体で 50epoch 学習.

学習条件として、バッチサイズは 24、最適化手 法は AdamW、学習率は 0.0001 で設定した. 検出 ヘッドに関するハイパーパラメータとして、Focal loss で  $\alpha = 0.25$ ,  $\gamma = 2.0$  とした. またセグメン テーションヘッドに関するハイパーパラメータと して、Tversky loss で  $\alpha = 0.7$ ,  $\beta = 0.3$ , Focal loss で  $\alpha = 0.25$ ,  $\gamma = 2.0$  を設定した. 全ての実験は NVIDIA RTX A6000 上で実行した.

# 4.1. 下流タスクにおける性能比較

**物体検出タスクの評価**. マルチタスク学習手法 である MultiNet, DLT-Net, シングルタスク学習 手法である Yolov5s, Faster R-CNN と性能を比較 する.本実験では, BDD100K における 10 クラ ス内で car, truck, bus, train クラスを vehicle クラスとし, vehicle クラスによる物体検出精度を 比較した.物体検出精度は AP50 と Recall で評価 する. AP50 とは,対象クラスに対する IoU の閾 値 0.5 における平均適合率である.

表1に物体検出タスクに対する評価を示す.提案 手法は Recall 94.9%, AP50 75.7% であり, Faster R-CNN, MultiNet, DLT-Net に対し精度が向上 し,高精度な物体検出モデル YOLOv5s と同等の 精度を獲得した.物体検出タスクにおける検出可 視化例を図3に示す.Faster R-CNN は車内のオ



Faster R-CNN

Ours

図3 物体検出タスクにおける検出可視化: 橙丸は未検出,赤丸は誤検出である.

**表**1 物体検出タスクに対する評価:\* はシングルタ スク手法である.

Method	Recall [%]	AP@.5 [%]	Speed[fps]
Multi-Net	81.3	60.2	8.6
DLT-Net	89.4	68.4	9.3
YOLOv5s*	86.8	77.2	82
Faster R-CNN <sup>*</sup>	77.2	55.6	8.8
Ours	95.0	76.5	31.0

ブジェクトや歩行者を車と誤認識しているが,提 案手法では未検出と誤検出が大幅に改善している.

**運転可能領域タスクの評価.** マルチタスク学習 手法である MultiNet, DLT-Net, シングルタスク 学習手法である PSP-Net [8] と性能を比較する. 本実験では, BDD100K データセット における direct, alternative クラスを drivable クラスと し, drivable クラスによるセグメンテーション精 度を比較した. 運転可能領域のセグメンテーショ ンタスクは mIoU で評価する.

表 1 に運転可能領域セグメンテーションタスク に対する評価を示す.提案手法は運転可能領域タ スクで mIoU 86.7% であり, MultiNet と DLT-Net を 15% 程度向上している. PSP-Net と比較し 2.9% 低下しているが,推論速度は3倍近く高速で あることが分かる.運転可能領域タスクにおける セグメンテーション可視化例を図 4 に示す.提案 **表 2** 運転可能領域に対する評価:\* はシングルタス ク手法である.

Method	mIoU [%]	Speed[fps]
Multi-Net	71.6	8.6
DLT-Net	71.3	9.3
$PSP-Net^*$	89.6	11.1
Ours	87.5	31.0

表3 車道線検出タスクに対する評価:\* はシングル タスク手法である.

Method	Accuracy [%]	IoU [%]	Speed[fps]
ENet*	34.1	14.6	100
SCNN*	35.8	15.8	19.8
ENet-SAD*	36.6	16.0	50.6
Ours	82.1	24	31.0

手法と PSP-Net のセグメンテーション結果は定性 的に同等の結果であることが分かる.これは,提 案手法では他タスクにより獲得したエッジ情報が セグメンテーションに貢献しているためであると 考えられる.

**車道線検出タスクの評価.**シングルタスク学習 手法である ENet [16], SCNN [17], ENet-SAD と 性能を比較する. BDD100K データセットの車道 線は,車道線に対するエッジとしてラベルが用意 されている.本実験では車道線の領域をセグメン テーションするため, BDD100K データセット内の



PSP-Net

Ours

図4 運転可能領域タスクにおけるセグメンテーション可視化

Object detection		Drivable area seg.	Lane line detection		
method	Recall [%]	mAP@.5 [%]	mIoU [%]	Acc [%]	IoU [%]
Ours	95.0	76.5	87.5	82.1	24.0
w/ CNN	94.8	75.5	85.8	79.1	21.3
w/o BiFPN	89.6	45.8	79.9	71.6	18.7

表 4 Next-ViT および BiFPN の有効性検証における下流タスクの精度比較

車道線ラベルを直接用いることはできない.そこ で、以下の手順で車道線のアノテーションを実施 した.まず、車道線に対するエッジを用いて中心 線を計算する.次に、テストデータでは中心線か ら線幅を2ピクセル、学習データでは線幅を8ピ クセルに設定し車道線ラベルとした.

車道線検出タスクに対する評価は Accuracy と IoU で評価する.表3に車道線検出タスクに対す る評価を示す.提案手法は Accuracy 80.7%, IoU 22.7% であり,この精度は従来のシングルタスク 手法を大きく上回っていることが分かる.車道線 検出タスクにおけるセグメンテーション可視化例 を図5に示す定性的結果から,提案手法は ENet-SADよりも明らかに車道線領域が正確かつ連続的 であることが確認できる.運転可能領域タスクと 特徴を共有することで,提案手法では運転可能領 域と車道線間での誤認識が大幅に抑制されている. しかし,未検出領域が存在し車道線検出領域はま だ不十分である.これは豊富なセマンティック情報を持つ低レベル特徴と高レベル特徴の融合方法 を改善することで,更なる精度向上が見込めると 考えられる.

## 4.2. Next-ViT の有効性評価

バックボーンネットワークとして用いた Next-ViT の有効性を検証するため,提案手法のバック ボーンネットワークを EfficientNet [18] とするモ デルと性能を比較する.下流タスクにおける定量 的評価を表 4,下流タスクにおける可視化例を図 6 に示す.ここで,w/CNN は EfficientNet を採用 したモデルを示している.図 6 から,提案手法は w/CNN より誤検出や未検出が少ない.また,表 4 から,提案手法は w/CNN より全下流タスクにお いて性能が向上している.したがって,Next-ViT は CNN ベースモデルと比較し特徴抽出能力が高 く,バックボーンネットワークに適していると言 える.



ENet-SAD

Ours

図 5 車道線検出タスクにおけるセグメンテーション可視化: 橙丸は未検出,赤丸は誤検出である.

#### 4.3. BiFPN の有効性評価

特徴融合手法として用いた BiFPN の有効性を 検証するため, BiFPN の有無によるモデルの性 能を比較する.下流タスクにおける定量的評価を 表 4,下流タスクにおける可視化例を図 6 に示す. ここで,w/o BiFPN は BiFPN を導入していな いモデルを示している.図 6 から,w/o BiFPN は多くの誤検出や未検出があることが分かる.ま た表 4 から,w/o BiFPN は提案手法と比較し,全 下流タスクに対してモデルの性能が大幅に減少し ている.したがって,下流タスクに合わせ複数レ ベルの特徴を融合することは,マルチタスク学習 において有効であると言える.

## 4.4. まとめ

4.1 節では,提案手法と従来のシングルおよびマ ルチタスク学習手法との性能比較について述べた. 実験から3つの異なる下流タスクに対して,従来 手法と同等もしくは精度向上を確認した.4.2,4.3 節では,提案手法における2つの要素であるNext-ViT と BiFPN の有効性を検証した.有効性検証 実験から,どちらも用いない場合では精度が大幅 に低下することを確認した.したがって,マルチ タスク学習においてバックボーンネットワークに Next-ViT,下流タスクへの特徴融合として BiFPN を導入することは,異なるタスク間での特徴抽出 を最適化し,モデルの性能向上に寄与すると考え られる.

#### 5. おわりに

マルチタスク学習では, バックボーンネットワー クの特徴抽出性能が下流タスクの精度に大きく影 響を及ぼす.また下流タスクごとに解く上で必要 な特徴レベルは異なり,下流タスクに合わせた特 徴が必要である.これら問題に対し,我々は自動運 転における Next-ViT と BiFPN を用いたマルチ タスクモデルを提案した.BDD100K データセッ トによる評価実験から,物体検出タスク,運転可 能領域タスク,車道線検出タスクの下流タスクに おいて高い性能を獲得することを確認した.しか し,従来手法と比較し本手法の推論速度は十分で あるとは言えない.そのため,今後はマルチタス クモデルにおける推論速度の高速化に取り組む予 定である.

## 参考文献

- Z. Zhang, et al., "Facial landmark detection by deep multi-task learning." ECCV, 2014.
- [2] S. Liu, et al., "End-to-end multi-task learning with attention." CVPR, 2019.
- [3] X. Zhao, et al., "A modulation module for multitask learning with applications in image retrieval." ECCV, 2018.



図 6 Next-ViT および BiFPN の有効性検証における可視化: 橙丸は未検出,赤丸は誤検出である.

- [4] M. Teichmann, et al., "Multinet: Real-time joint semantic reasoning for autonomous driving." IV, 2018.
- [5] Y. Qian, et al., "DLT-Net: Joint detection of drivable areas, lane lines, and traffic objects." IEEE Transactions on intelligent transportation systems, 2019.
- [6] A. Bochkovskiy, et al., "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934, 2020.
- [7] S. Ren, et al., "Faster r-cnn: Towards real-time object detection with region proposal networks." NeurIPS, 2015.
- [8] H. Zhao, et al., "Pyramid scene parsing network." CVPR, 2017.
- [9] J. Li, et al., "Next-ViT: Next Generation Vision Transformer for Efficient Deployment in Realistic Industrial Scenarios", arXiv preprint arXiv:2207.05501, 2022.
- [10] M. Tan, et al., "EfficientDet: Scalable and Efficient Object Detection", CVPR, 2020.
- [11] A. Dosovitskiy, et al., "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929, 2020.
- [12] T. Lin, et al., "Focal Loss for Dense Object Detection", ICCV, 2017.
- [13] S. Salehi, et al., "Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks", MLMI, 2017.
- [14] F. Yu, et al., "Bdd100k: A diverse driving dataset for heterogeneous multitask learning." CVPR, 2020.
- [15] Y. Hou, et al., "Learning lightweight lane detection cnns by self attention distillation." ICCV, 2019.
- [16] A. Paszke, et al., "ENet: A deep neural network architecture for real-time semantic segmentation." arXiv preprint arXiv:1606.02147, 2016.
- [17] X. Pan, et al., "Spatial as deep: Spatial cnn for traffic scene understanding." AAAI, 2018.
- [18] M. Tan, et al., "Efficientnet: Rethinking model scaling for convolutional neural networks." ICML, 2019.

張 陳雨:2020年同済大学浙江学院電子情報工学科卒業.現 在中部大学大学院修士課程在学中,自動運転においてマルチ タスク学習の研究に従事. 平川 翼:2017年広島大学大学院博士課程後期課程情報工学 専攻修了.2017~2019年中部大学研究員,2019年同大学特 任助教,2021年同大学講師.2014年独立行政法人日本学術 振興会特別研究員 DC1.2014~2015年ESIEE Paris 客員 研究員.博士 (工学).

山下 隆義:20002 年奈良先端科学技術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,2011 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 年同大学准教授,2021 年同大学教授.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. 博士 (工学).

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年同大准教授 を経て2010年より同大教授.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.博士(工学).