# 明日のトップランナー

NTTコミュニケーション科学基礎研究所

## **金子卓弘** 特別研究員

### 二次元画像から三次元情報を推定 AR-GANで「三次元世界を理解する」コンピュータの実現へ

**ヒトは写真を見れば，これまでに培った経験や知識などから奥行きなどを推定することができますが，コンピュータにとっては容易なことではありません．今回は，無作為に収集した一般的な画像から三次元情報を学習可能な深層学習モデルを構築した金子卓弘特別研究員にお話をお聞きしました．**

◆PROFILE：2014年東京大学大学院修士課程修了．同年，日本電信電話株式会社に入社，NTTコミュニケーション科学基礎研究所に所属．2020年東京大学大学院博士課程修了．博士（情報理工学）．2020年よりNTTコミュニケーション科学基礎研究所特別研究員．画像生成，音声合成，音声変換を対象としたコンピュータビジョン，信号処理，機械学習，深層学習の研究に従事．日本機械学会畠山賞，ICPR Best Student Paper Award，音声研究会研究奨励賞，東京大学大学院研究科長賞等を各受賞．
詳細はこちら：https://www.kecl.ntt.co.jp/people/kaneko.takuhiro/

## AR-GANによる奥行きとボケ効果の学習

**◆「一般的な画像から奥行きとボケ効果を学習する」とはどのような研究なのでしょうか．**

私たちの生活している三次元世界をそのまま記録することは困難です．そのため三次元情報の代わりに写真などの二次元画像を記録・保存することがよく行われています．

ヒトは写真を見れば，これまでの経験や知識などから奥行きなどの三次元情報を推定することができますが，コンピュータにはそうした経験や知識がないため，写真から三次元情報を推定することは簡単なことではありません．今後，ロボットが私たちの生活をサポートするような場面を考えたとき，コンピュータが三次元情報を理解することは必要不可欠となります．コンピュータにとってもっとも学習しやすい方法は，学習データとして二次元画像と三次元情報のペアを大量に与えることです．正解が分かっているので学習も容易といえます．しかし，この方法では測距センサや2つのカメラを搭載したステレオカメラなどの特殊な機器が必要となり，コストもかかります．

そのため，本研究では普通のカメラで撮影した身近な画像や，Web上で見かけるようなごく一般的な画像から三次元情報を学習可能な深層学習モデルを構築しました．撮影した写真を見ると，対象物にはピントが合っていて，背景はボケていることが多いと思います．この「ボケ効果」を手掛かりとして，三次元情報，なかでも特に奥行き情報や，どんなボケ効果が付いているかなどを学習します．言い換えれば，三次元情報をカメラを通して二次元画像に射影する問題を順問題とすると，本研究は，いろいろな情報が欠落している二次元画像のみから三次元情報を推定するという逆問題，いわゆる「不良設定問題」を解こうという取り組みです．

**◆具体的にはどのような仕組みで学習が行われているのでしょうか．**

本研究はGAN（Generative Adversarial Network：敵対的生成ネットワーク）と呼ばれる技術に基づいています．GANは正解があらかじめ決められていない，いわゆる「教師なし学習」の一種で，生成器と識別器の2つのニューラルネットワークで構成されています．生成器は乱数を使用していわば「偽物の画像」を生成する働きを持ちます．一方，識別器は「本物の画像」と生成器が生成した「偽物の画像」の2種類を使用し，本物か偽物かを見分ける働きを持ちます．生成器はなんとか識別器をだまそうとし，識別器はなんとか正確に識別しようと敵対するため，競争しながら学習を進められ，結果として生成器はリアリティのある画像を生成できるようになるという仕組みです．

GANは二次元の画像を生成することに特化した技術ですので，三次元世界との結びつきは持ちません．そこで，GANにカメラの光学的な性質を組み込んだ「AR-GAN（Aperture Rendering GAN）」を提案しました．「Aperture」は日本語で

はカメラの「絞り」にあたります．カメラが三次元世界を二次元画像に射影する際の絞りによる光学的な制約を組み込むことで，生成器は二次元画像，奥行き情報，ボケ効果の３つを関連付けながら学習することができます．

**図**はAR-GANの生成器における処理の流れを示したものです．「画像生成器」はGANにも含まれるもので，乱数が与えられると画像を生成します．一方，奥行き生成器は画像と対になる奥行き情報を生成するもので，AR-GAN独自のものです．続いて生成されたこれらの一対のデータを使用してカメラの絞りを模擬した機構を実現します．光線場は25枚の画像で構成されており，カメラの絞りの中にどのような画像が映っているかを表しています．中央には絞りの中心を通って入ってきた光による画像を表しています．中央から右にずれたところには絞りの右部から対象物を見た画像，上にずれたところには絞りの上部から対象物を見た画像が表示されます．

この機構は顔を上下左右に動かしながら目の前にある対象物を見るときをイメージしていただくと分かりやすいかもしれません．ピントが合っている対象物は顔を動かしても同じ位置にあり動かないため鮮明に写りますが，それより遠くにあるものは顔を動かすと大きく動きます．そのためこれらの画像を足し合わせるとボケのある画像が生成されるわけです．このようにしてカメラの絞りの違いによる写り方の違いを表現する機構となっています．

**◆現在までの研究の進捗と課題について教えてください．**

現在は花画像，鳥画像，ヒトの顔の画像などが実際に生成できるようになっています．学習には数日かかりますが，学習が完了していれば数秒程度で本物の画像と区別ができないようなレベルのボケのある画像を生成することができます．現時点では対象の種類を絞れば生成できそうだ，という感触を持っています．

課題としては，画像の種類により学習しやすいものと学習しにくいものがあることがあげられます．例えばヒトの顔画像の場合，パーツの形状や位置がある程度決まっているため学習しやすく，学習しにくいものとしてはさまざまな視点や距離で撮影した動物の画像などバリエーションの多いものがあげられます．

また，画像サイズが大きくなればなるほど，より細部まで表現する必要があるため，処理時間が増加し，また，学習も難しくなります．今後適用範囲を広げていくにあたり，いろいろな課題が出てくるのではないかと思います．やはり不良設定問題特有の難しさを感じますね．

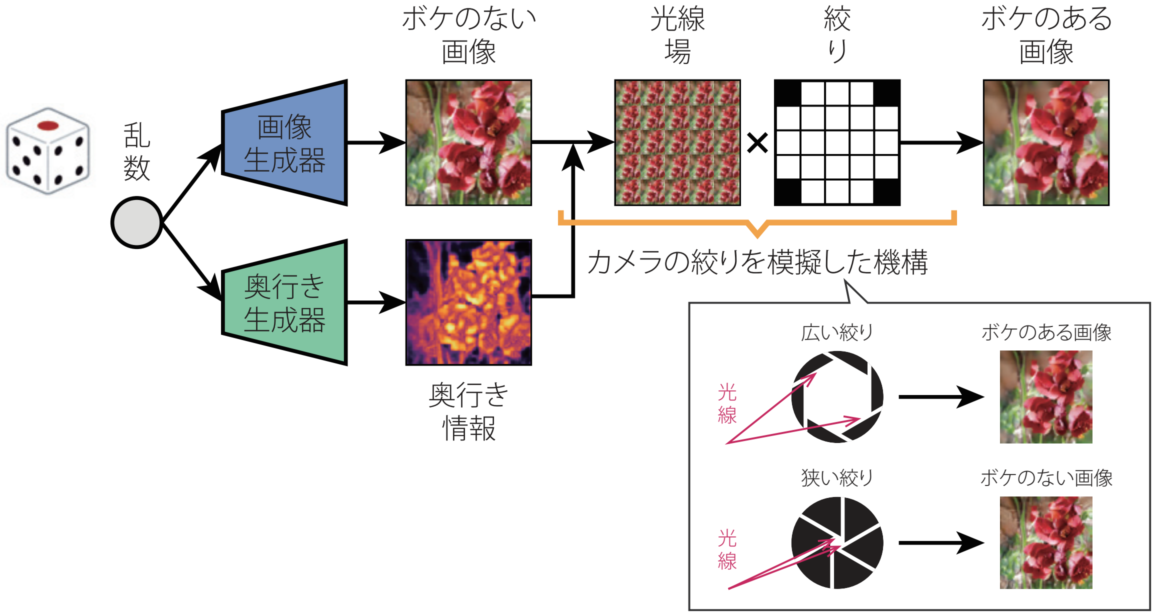

図　AR-GANの生成器における処理の流れ

## 三次元世界を理解するコンピュータを実現

**◆本技術により，将来どのようなことが可能となるのでしょうか．**

私たち研究者のミッションとして，「ヒトと親和性の高いコンピュータの実現」があります．そのために三次元情報の理解は必要不可欠だと思いますが，より幅広い分野に応用するためにはデータ収集コストが障壁になりそうです．その点で，データ収集に優れた本研究は有益ではないでしょうか．将来的には三次元世界を自由に動き回れるロボットを実現したり，仮想空間上に違和感のない三次元世界を構築したり，仮想空間で三次元物体を制作するツールを開発したりなどの活用が見込めます．

また，二次元画像さえ集めればそれにフィットしたモデルを構築できるということも利点です．例えば高名な写真家が撮影した写真を集めれば，その写真家特有のボケ効果を学習したモデルを構築できます．現在はSNSなど画像を使ったコミュニケーションが活発に行われています．ボケ効果を手軽に利用できるようになれば，より「映える」写真も簡単につくれるようになるかもしれません．

特にロボティクス分野，コンテンツ生成分野，エンタテインメント分野の３分野に対してはかなり有用だと思います．

**◆今後の展開，他分野とのコラボレーションについて教えてください．**

基礎研究なので，現時点では実用化に向けた具体的な目標を設定することは難しいですが，今後も高精度化，高解像度化などで性能を向上させていくということは考えています．今回はカメラの絞りでしたが，物理的な制約を入れることでより信頼性の高いコンピュータが創り出せるという点は面白いですね．


（今回はリモートにてインタビューを実施しました）

IOWN（Innovative Optical and Wireless Network）構想の主要技術分野の1つに，多様な産業やモノとヒトのデジタルツインを自在に掛け合わせて演算を行う「デジタルツインコンピューティング（DTC）」があります．実世界と仮想世界とを融合させるにはコンピュータが実世界をしっかりと理解できることが必要となりますので，そういった分野にも寄与できるかなと考えています．

私が手掛けるメディア生成技術という分野については，技術の成熟に伴ってさまざまな分野との融合が重要度を増していると感じています．現在はコンピュータビジョンや機械学習などのコンピュータ科学と，光学などの物理学を組み合わせた研究をしていますが，今後は画像作成に関してはコンピュータグラフィックス分野，写真撮影に関してはフォトニクス分野など，異分野の方との交流や異分野の導入にも力を入れていきたいと思います．

**◆若手研究者，および将来のビジネスパートナー様に向けてメッセージをお願いいたします．**

NTT研究所では基礎研究から応用まで広範囲に研究しており，特に，私の所属するNTTコミュニケーション科学基礎研究所では，ヒトとヒト，ヒトと機械のコミュニケーションをいかに良くしていくかという研究を行っています．これまでは基礎研究にとどまることも多かったのですが，最近では基礎研究と応用研究の距離が縮まってきており，現実的な問題を考えながら研究に取り組む側面が増加していると感じています．国際会議で発表された技術がアプリケーションに搭載されたり，Web上のサービスとして展開されたりというケースも増えてきました．数式をいじって終わりではなく，例えば実際に声を変換するソリューションを創り出すなど，アウトプットが見えるかたちとなる点は面白いのではないでしょうか．

個人での研究には限界があります．当研究所では大学との共同研究やインターンシップ生の受入れなども活発に行っていますので，特に何かをつくりたい，何かを変えたいと思っている学生の皆さんや若手研究者の皆さんとは今後も積極的にコラボレーションしていきたいな，と思っています．

ビジネスパートナーの皆さんに関しても，研究者発信で面白いものができたから是非サービスに結び付けてほしいとお願いすることもありますし，逆に実問題に造詣が深いサービス分野の方からフィードバックをいただいて研究のアイデアを得ることもありますので，今後も活発に交流していきたいと考えています．