Tokyo Tech

2022 年 5 月 12 日

報道機関各位

東京工業大学

機械学習の進化が、「レンズ」というカメラの当たり前を覆す

－次世代イメージセンシング・ソリューション開発を加速－

【要点】

○ 最先端機械学習モデル「Vision Transformer」に基づく、新たなレンズレスカメラの画像再構成手法を提案

○ 提案した画像処理技術は高速に高品質な画像を生成できることを実証

○ 小型・低コストかつ高機能であるため、IoT 向け画像センシング等への活用に期待

【概要】

東京工業大学 工学院 情報通信系の潘秀曦（Pan Xiuxi）大学院生（博士後期課程 3 年）、陈啸（Chen Xiao）大学院生（博士後期課程 2 年）、武山彩織助教、山口雅浩教授らは、レンズレスカメラの画像処理を高速化し、高品質な画像を取得できる、**Vision Transformer**（**ViT**、用語 1）と呼ばれる最先端の機械学習技術を用いた新たな画像再構成手法を開発した。

カメラは通常、焦点の合った画像を撮影するためにレンズを必要とする。現在、**IoT**（用語 2）の普及に伴い、場所を選ばず設置できるコンパクトで高機能な次世代カメラが求められているが、従来のカメラは、レンズ系のサイズのために極端な小型化が難しい。これに対して、コンピュータを用いた画像再構成処理の応用によって、レンズを使用せずに画像を取得する「レンズレスカメラ」を実現することが提案され、注目を集めている。しかし、これまでに、その画像再構成の技術が確立されておらず、画質が不十分で計算時間がかかるなど実用面での制約があった。今回開発した手法は、ViT を用いることで従来の**畳み込みニューラルネットワーク（CNN）**（用語 3）に基づく深層学習の課題を解決し、短時間の計算で高品質な画像を出力できる。本技術は、超薄型・軽量・低コストのレンズレスカメラの実用性を大きく高めると期待される。

本研究成果は、3 月 31 日付の科学誌「*Optics Letters*」に公開された。

●背景

近年、“レンズレスカメラ”と呼ばれるデバイスに注目が集まっており、その名の通り、レンズのないカメラを指している。従来のカメラは、被写体から発せられた光をレンズ系でイメージセンサー上に集光することによって画像を撮影しているが、高画質で明るく劣化のない画像を得るためには、複雑なレンズ系が必要となりカメラが大型化する要因となっている。一方、近年の IoT の普及等に伴い、カメラの小型化、軽量化、低価格化への要求が高まっている。それらの要求に応えるのがレンズレスカメラである。

レンズレスカメラの基盤となっているのは、近年の計算技術の進歩である。特に、画像撮影過程の一部を光学ハードウェアからコンピューティング技術に置き換えることで、レンズ光学系を簡素化する技術が研究されている。レンズレスカメラにおいては、被写体から発せられた光がシート状の特殊なマスクを通して符号化され、センサー上にパターンとして投影される。このパターンは人間の目には全く理解できないが、光学系の特性を数値的にモデル化することで復号（画像として再構成）できるようになる。数学的処理の手法を変えることで、撮影後にピントを自在に変える**リフォーカス**（用語 4）などができるのも、従来のカメラにはない特徴である。究極的には、マスクとセンサーを半導体プロセスで一体的に作製することも可能である。また、レンズの制約がなくなるので、超小型のデバイスで今までにない新しいアプリケーションを可能にすると期待される（図 1、2）。

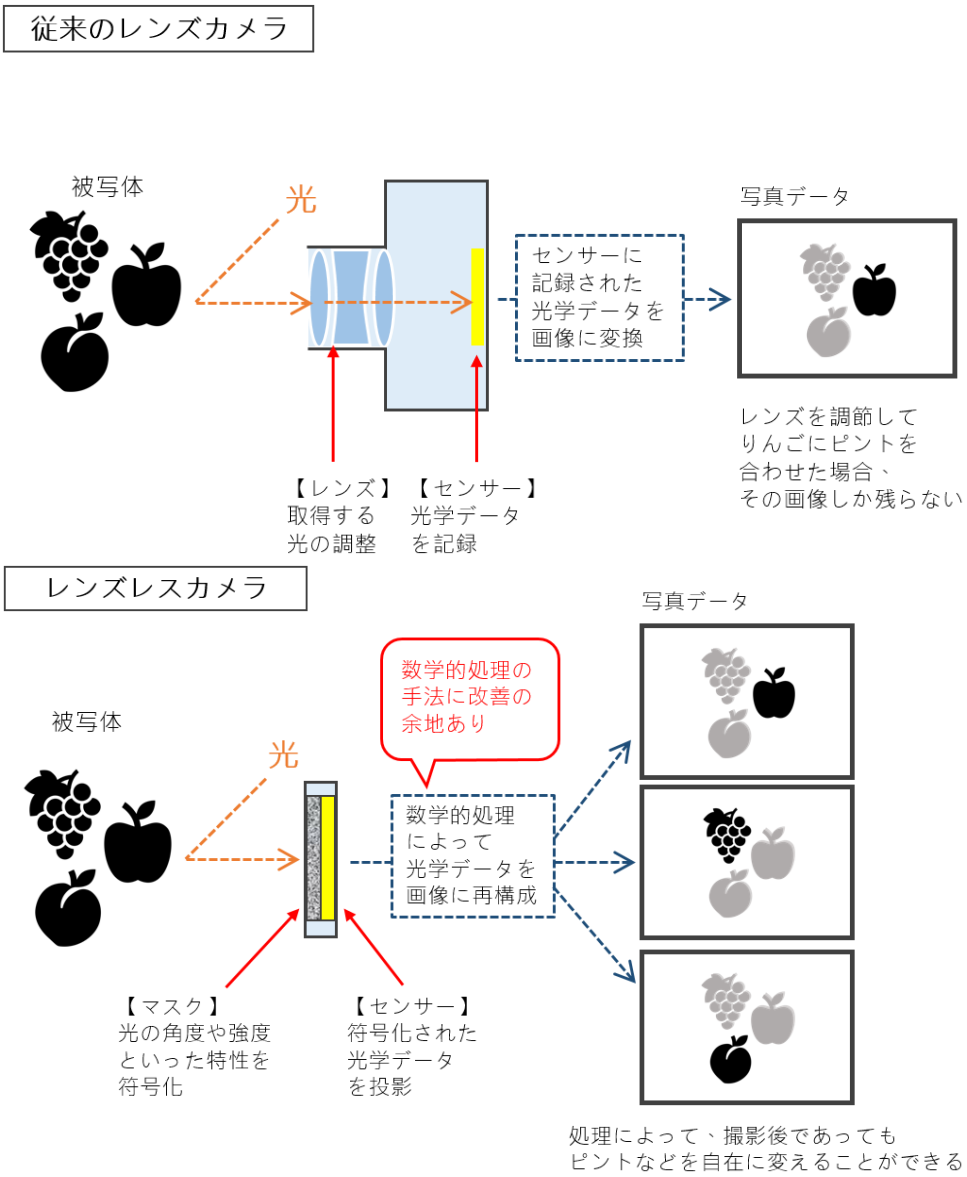

図 1. 従来のカメラとレンズレスカメラのしくみ。

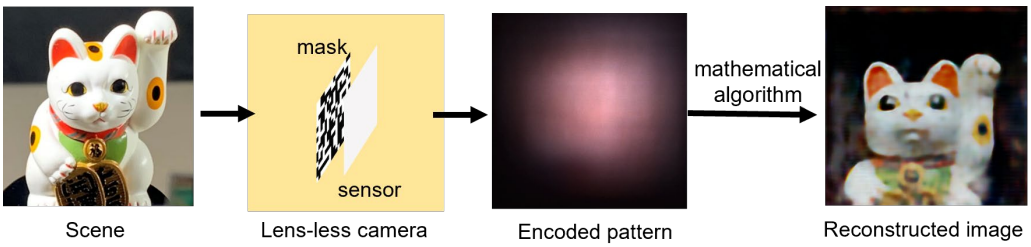

図 2. レンズレス撮影の流れ。マスクを通してイメージセンサー上にパターンを投影する。その後、数学的なアルゴリズムによって画像を再構成する。

しかし、レンズレスカメラ技術の実用化に向けては、画像再構成技術に基づく復号処理に課題が残されている。これまで用いられてきた復号処理には、①モデ

ルに基づく復号法、ならびに②機械学習を用いた手法がある。まず、従来のモデルに基づく復号法では、レンズレス光学系の物理モデルを数学的に近似し、**凸最適化問題**（用語 5）を解くことで画像を再構成するが、物理モデルを精度よく近似することが難しく、画像の品質が低下しやすい。また、最適化問題を解くアルゴリズムは通常反復的な計算が必要となり、処理時間が長くかかってしまう。ディープラーニングのような機械学習を用いると、物理モデル近似のエラーや処理時間の問題を回避することができるが、画像再構成に対するディープラーニング手法として一般的に採用されている畳み込みニューラルネットワーク（CNN）は、レンズレス光学系の特性に適していないという課題があった。CNN は隣接した画素同士の相互関係を主に学習するのに対して、レンズレス光学系では被写体の一点からの情報をイメージセンサー上の広い範囲に投影するので、画像上の広い範囲の相互関係を学習しなければならない。このため、効率的な処理ができず、十分に高品質な画像が得られなかった。

●研究成果

本研究では、マスクを用いたレンズレス光学系の物理モデルとその特性の考察に基づき、画像再構成のための新しい機械学習アルゴリズムを提案した（図 3）。提案するアルゴリズムは、近年注目されている ViT と呼ばれる最先端の機械学習技術に基づいている。図 3 のアルゴリズムでは、重なり合いを持つパッチ化モジュールを用いた多段階のトランスフォーマーブロックの構造に新規性がある。これによって画像の特徴を階層的な表現で効率的に学習することができる。

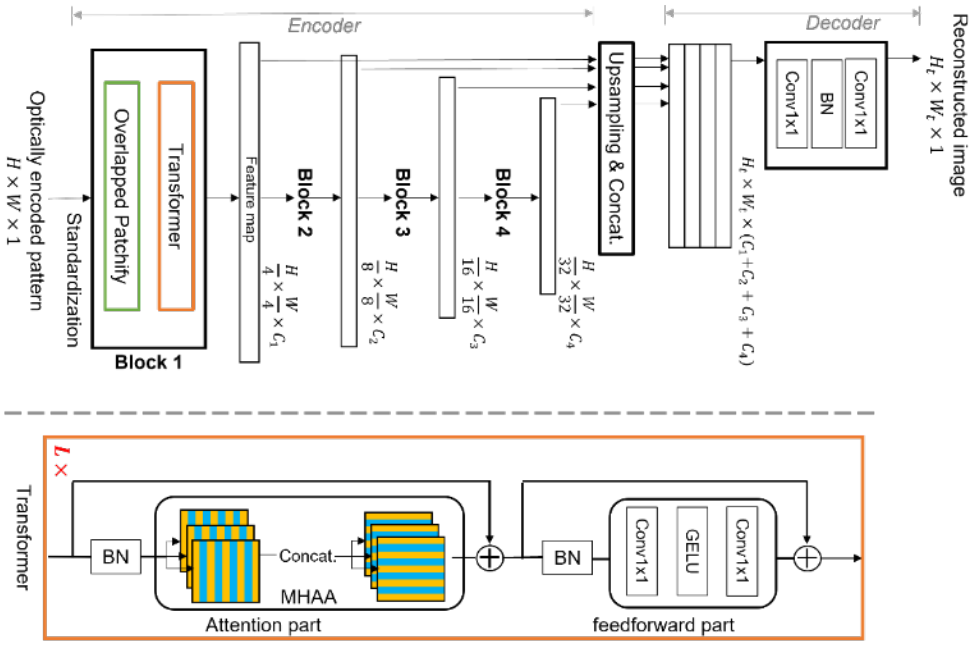

図 3. ViT を基にして考案された画像再構成のための深層ニューラルネットワークの構造

従来のモデルに基づく手法は反復的な処理のために計算時間を要するのに対して、提案手法は機械学習を用いた反復不要な処理アルゴリズムであるため、高速な処理が可能である。また、物理モデルをコンピュータが学習するため、モデル近似誤差の影響を格段に減らすことができる。加えて、従来の機械学習を用いた復号手法では CNN が画像内の局所的関係を主に学習していたのに対して、提案手法は画像内の大局的な特徴量を利用するため、イメージセンサーの広い範囲にわたる投影パターンの処理に適している。

以上のように、提案手法では従来手法が抱えていた処理時間および品質の限界を ViT アーキテクチャによって解決し、短い演算時間で高品質な画像を取得することが可能になった。

図 4 に示す装置を用いた光学実験の結果、提案した再構成法を用いたレンズレスカメラは、図 5 のように他の従来の手法よりもノイズが少なく鮮明な画像を生成できた。さらに、処理の計算速度が十分速く、リアルタイムでの撮影も可能である。

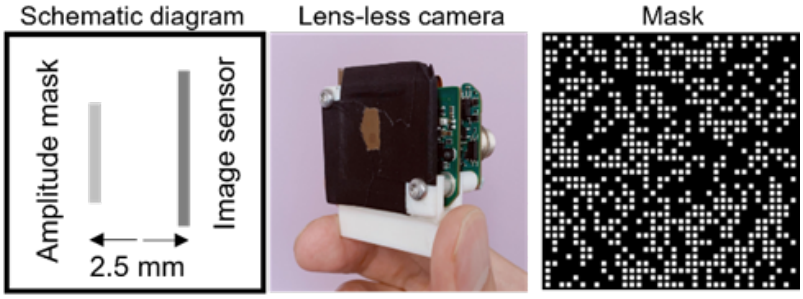

図 4. 光学実験に使用したレンズレスカメラ。レンズレスカメラは、マスクと 2.5 mm 離れた位置のイメージセンサーで構成されている。マスクは、開口サイズ 40×40 μm の合成石英板にクロムを蒸着して作製した。

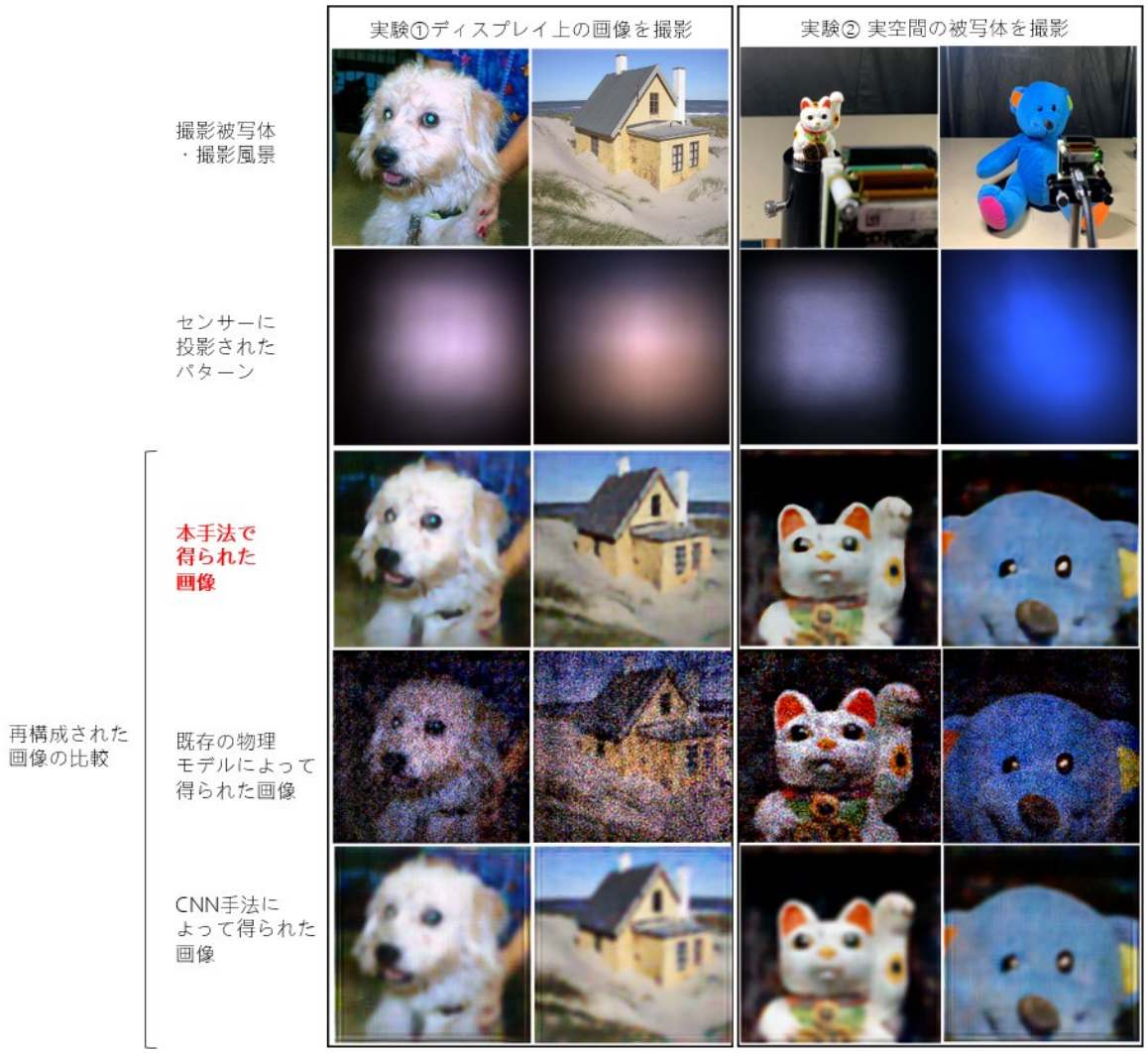

図 5. 光学実験結果。撮影対象は、液晶画面に表示された画像（左 2 列）と、実物体（招き猫人形とぬいぐるみ、右 2 列）である。1 行目は、液晶画面に表示された原画像と、実物体の撮影風景を示したものである。2 行目はセンサーに撮影されたパターンを示している。最後の 3 行は、提案手法、モデルベース手法、CNN ベース手法による再構成画像をそれぞれ示している。提案手法は、最も高品質でノイズが少なく鮮明な画像を生成できている。

●今後の展開

レンズレスカメラの利点は小型化だけではなく、レンズの製造が難しい不可視光イメージングへの適用や、マスクを通じて撮影した画像からワンショットでの 3D 撮影への応用などにも発展可能である。レンズレスカメラは、小型でありながら高機能という特徴を持つことから、次世代画像センシングソリューションの新たな方向性として、これからの IoT の進化を支えることが期待できる。

【参考情報】
山口雅浩研究室ホームページ「コンピュテーショナルイメージングによるレンズレスカメラ・プロジェクター技術」
https://www.oid.ict.e.titech.ac.jp/cn2/pg95.html

【付記】
本研究は超スマート社会リーダーシップ博士奨励金、JST 次世代研究者挑戦的研究プログラム JPMJSP2106 の支援を受けて行われた。実験の一部は、中村友哉元東京工業大学助教（現大阪大学准教授）の協力により行われた。

【用語説明】
（1） **Vision Transformer**：2020 年に Google が発表したディープラーニングによる画像認識モデル。自然言語解析の分野で開発された Self-Attention 構造に基づく Transformer を画像認識へ応用したもので、従来とは大きく異なる構造で高い精度を達成したことから注目されている。
（2） **IoT**：Internet of Things（モノのインターネット）の略語で、実世界に存在している様々なものをインターネットに接続し、ネットワークを通じて相互に情報交換を行う仕組みのこと。人々の生活のあらゆる場面を変革する技術として実用化が進んでいる。
（3） **畳み込みニューラルネットワーク（Convolutional Neural Network: CNN）**：ディープラーニングで広く使われる技術の一つで、その高性能化へ大きく寄与し、コンピュータビジョンや音声認識、その他の様々な分野を大きく発展させた。画像や信号のあるサンプル点と近くのサンプル点の位置関係だけでネットワークの接続重みが定まるので、効率的に学習や処理が行える。
（4） **リフォーカス**：通常の写真撮影では撮影した際にピントを合わせた位置が鮮明に写り、ピントの合っていない箇所はぼけてしまうが、撮影後の写真から、画像処理によってピントを前後に移動させたり、被写界深度を変えたりすることをリフォーカスという。撮影時に特殊な光学系を用い、撮影後にコンピュータを用いた画像処理を適用することによってリフォーカスを実現する方法が開発されている。
（5） **凸最適化問題**：凸集合上の凸関数の値を最小化するような変数の値を求める数学的な問題である。例えば画像再構成の問題では、撮影されたセンサー上のデータから最適化問題を解くことによって元の物体の情報を求めることができる。凸最適化問題は局所的な最小値が大域的な最小値と一致する性質をもつため、一般的な最適化問題よりも容易に計算することができる。

【論文情報】
掲載誌： *Optics Letters*
論文タイトル：Image reconstruction with transformer for mask-based lensless imaging
著者：Xiuxi Pan, Xiao Chen, Saori Takeyama, Masahiro Yamaguchi
DOI：10.1364/OL.455378

【問い合わせ先】
東京工業大学 工学院 情報通信系
教授 山口雅浩
E-mail：yamaguchi.m.aa@m.titech.ac.jp
Tel：045-924-5137

【取材申し込み先】
東京工業大学 総務部 広報課
Email: media@jim.titech.ac.jp
TEL: 03-5734-2975　FAX: 03-5734-3661