



Deep Learning Approaches to High-Speed Imaging in Constrained Environmental and Computational Conditions

庄, 緯豪

(Degree)

博士 (工学)

(Date of Degree)

2024-03-25

(Date of Publication)

2026-03-25

(Resource Type)

doctoral thesis

(Report Number)

甲第8944号

(URL)

<https://hdl.handle.net/20.500.14094/0100490169>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



(別紙様式 3)

論文内容の要旨

氏 名 _____ 庄 緯豪 ZHUANG WEIHAO _____

専 攻 _____ 情報科学 _____

論文題目 (外国語の場合は, その和訳を併記すること。)

Deep Learning Approaches to High-Speed Imaging in
Constrained Environmental and Computational
Conditions

(制約された環境および計算条件におけるハイスピードイメ
ージングのための深層学習アプローチ)

指導教員 _____ 滝口 哲也 _____

Chapter 1: Introduction

How fast can we see? The human visual system is estimated to perceive the world at a rate of 30 to 60 FPS. Any movement or oscillation faster than this frequency is imperceptible to our naked eye. Augmented with cameras and computers, we are now able to record images of the world beyond hundreds of thousands of FPS, allowing us to observe phenomena that cannot be captured by the naked eye.

The ability to make near instantaneous observations of the world is key to many scientific research endeavors. For example, rocket combustion experiments in the aerospace research rely on high-speed imaging devices for data collection, as a minimum of 100,000 FPS is required to analyze the extremely fast movements resulting from combustion. Image cytometry in biology also relies on high-speed cameras to record the characteristics of each cell among tens of thousands of cells, making it convenient for researchers to filter out the cells useful for research. Fluid dynamics, structure mechanics, and many other fields similarly rely on the ability to perceive the world at high frame rates. Beyond the sciences, industrial processes often rely on high-speed imaging for both quality control on production lines and development purposes.

In the past decades, high-speed camera manufacturers have been pushing the limits of high-speed imaging, enabling the study of ever faster and higher frequency phenomenon. Early high-speed imaging relied on rotating mirrors film cameras, which developed from cameras designed by Kodak that could capture images at a rate of 5,000 FPS, and progressed to technology capable of capturing 1M FPS. Subsequently, although there were a series of different types of high-speed film cameras (such as streak cameras and rotating prism cameras), film-based high-speed cameras have reached their theoretical maximum performance due to physical limitations.

The emergence of digital imaging technology, especially the use of silicon-based CCD and CMOS sensors, became the mainstream development approach for high-speed camera sensors by achieving higher frame rates, lower power consumption, and on-chip processing. CCD structures can provide high-quality and low-noise images, but CMOS can parallelize and partially read the subregions of sensors at high rates. Due to the widespread use of CMOS structures in consumer devices, they have received more development, and currently, many commercial companies (e.g., Photron, Vision

Research, DEL Imaging, IX Cameras, NAC, Weinberger, Shimadzu and LaVisionn) are working on developing CMOS-based high-speed cameras. In the early stages, Etoh et.al. developed a high-speed camera based on CMOS sensors that could capture images at 4500 FPS. Recent advancements have enabled high-speed cameras to achieve 100M FPS.

In digital imaging, the endeavor to faithfully capture and store visual data from a scene is a sophisticated procedure that involves several integral stages. The overarching objective is to encapsulate the essence of the scene within the confines of a digital device. The initial stage involves harnessing the raw visual data from the scene through a sensor. This sensor transforms the analog scene information into a discrete digital signal, typically referred to as a 'RAW' image. An intrinsic characteristic of this transformation is the introduction of noise, an artifact that often becomes more prominent with increased shutter speed. Subsequent to this initial acquisition, the RAW image undergoes further refinement. Typically, an Image Signal Processor (ISP) within the imaging device processes this RAW data, optimizing it to produce an RGB image. This RGB image is a representation closer to human visual perception, encapsulating the richness of colors and details inherent in the scene. Once processed to satisfaction, the RGB image is then committed to the device's storage, ensuring that the visual data is preserved for future access and analysis.

However, extreme observation rates are challenging for the following two fundamental limits:

1. The amount of light recorded by the camera sensor is inversely proportional to the rate of acquisition. For extremely high recording speed, the signal to noise ratio of recorded light becomes too small for traditional ISP to process.
2. The amount of data transfer, processing, and storage capacity needed grows proportional to the recording rate. Thus, the speed at which one can record is fundamentally limited by the available hardware and software.

In this dissertation, we mainly introduce two optimization algorithms designed to address limitations encountered by imaging devices during imaging. The first approach aims to address underexposed images arising from insufficient ambient lighting by deploying image denoising techniques. The second strategy seeks to overcome transmission limitations of imaging devices by employing video frame interpolation,

thereby achieving higher FPS. In addition, we propose a method intended to minimize the memory consumption of deep learning (DL) models, facilitating more efficient deployment to hardware devices.

Chapter 2: Optimization of Imaging Quality

In this chapter, we presented a novel approach to enhance the capabilities of high-speed cameras by treating the problem of high-speed imaging as an underexposed image denoising problem. Our method enables high-speed cameras to record clear images beyond one order of magnitude faster shutter speed without relying on artificial light sources. This advancement significantly expands the range of phenomena that can be studied using high-speed imaging, opening new doors for research in various scientific and industrial fields. We combining RNS, UPI and InvISP address the practical difficulty of data collection with non-portable high-speed cameras by requiring only the collection of bias frames from the target device and leveraging large public datasets to synthesize a diverse set of training data. This approach not only minimizes the data collection efforts but also enhances the applicability of our method to a wider range of high-speed cameras and experimental setups. To validate the effectiveness of our method, we use HSC collect a dataset of static and dynamic scenes. The quantitative evaluation on static scenes demonstrated the accuracy of our denoising results, while the qualitative assessment on dynamic scenes illustrated the improvements in imaging fast-moving phenomena. These results showcase the potential of our method in significantly advancing the capabilities of high-speed cameras, which is valuable for research areas that depend on capturing fast-moving events.

Chapter 3: Optimization of FPS for Imaging Devices

In this chapter, we have shown that SIREN representations of videos can be constrained to satisfy the OF constraint equation in their exact derivatives. We have seen that OF-constrained SIREN reach state of the art VFI on limited motion ranges, without relying on ML based residual flow and interpolation. We have also shown that the OF constraint not only allows SIREN to generate intermediate frames, but can also improve the ability of narrow SIREN to fit observed frames. We have discussed the limitations of our approach in its current form and outlined potentially impactful way forwards for future research.

Chapter 4: Memory Optimization for Deep Learning Model Deployment in Imaging Devices

In this chapter, we propose a method to reduce the memory consumption of convolutional neural networks in the inference phase that is orthogonal to other model compression methods. Using our method and the analytical results of this formulation, we can design a more memory-friendly model. We have used our proposed method in some successful CNN models, resulting in a reduction in the memory consumption of the model in the device. However, our proposed method also imposes a certain computational overhead, showing that it is more effective for high-resolution input images.

Chapter 5: Conclusions

In this chapter, we draw the conclusions for this dissertation.

氏名	庄 緯豪		
論文 題目	Deep Learning Approaches to High-Speed Imaging in Constrained Environmental and Computational Conditions (制約された環境および計算条件におけるハイスピードイメージングのための深層学習アプローチ)		
審査委員	区 分	職 名	氏 名
	主 査	教授	滝口 哲也
	副 査	教授	太田 能
	副 査	教授	的場 修
	副 査	准教授	高島 遼一
	副 査		
要 旨			
本研究では、高フレームレートで映像を記録するハイスピードイメージングの技術について、撮影・記録・実装の観点から三つの改善方法を提案している。人間の視覚システムにおいて情報を捉えられる速度はカメラのフレームレートで例えると 30 から 60FPS 程度と言われている。肉眼では知覚できない高速な現象を、数十万 FPS を超える速度で記録して観察することを可能とするハイスピードイメージングシステムは、航空宇宙研究、生物学の画像サイトメトリー、流体力学、構造力学など多くの科学分野において重要視されている。また、工業の分野においても、生産ラインの品質管理や開発目的で頻繁に利用されている。 ハイスピードカメラの開発は、現在は 100 万 FPS 程度のデジタル技術へ進化しているが、さまざまな環境、計算資源での撮影を可能にするためには、以下の二つの根本的な制限が考えられる。 1. カメラセンサーによって記録される光の量は、取得速度に反比例する。非常に高速で記録する場合、記録された光の信号対ノイズ比が小さすぎるため、従来の信号処理では明瞭な画像を生成できない。 2. データ転送、処理、およびストレージ容量の必要量は、記録速度に比例して増加する。したがって、記録できる速度は、利用可能なハードウェアとソフトウェアによって根本的に制限される。 本研究では、イメージングシステムが撮影中に直面する上記の制限に対処するために、三つの改善手法を提案している。一つ目の提案手法として、不十分な光源が原因で不明瞭な画像が生成される問題を解決するため、画像のノイズ除去を行う手法を提案している。二つ目の提案手法として、動画再生時にフレーム間の情報を補間することで、記録するフレーム数を減らしても高フレームレートで再生可能にする手法を提案している。この手法により、記録するフレーム数を減らせるため、（空いたストレージに）解像度の高い画像を記録することが可能となる。しかし近年主流となりつつある深層学習モデルベースの画像処理システムでは、高解像度画像の処理時に深刻なメモリ負担を引き起こすという新たな課題が生じることとなる。そこで三つ目の提案手法として高解像度の画像を処理する際の深層学習モデルのメモリ消費を減少させる方法を提案している。 第一章では序論として、ハイスピードイメージング技術の重要性と課題、および本論文の構成について述べている。 第二章では、不十分な光源が原因で不明瞭な画像が生成される問題を解決するため、画像のノイズ除去について提案している。速い動きの現象を記録する際、長時間露光による被写体のブレと短時間露光によるアンダーエクスポージャーノイズのトレードオフが存在する。高速撮影において、カメラの撮像素子が受け取る光は露光時間の短縮に伴って減少するため、短時間露光ではアンダーエクスポージャーノイズと呼ばれる、光の不足に起因するノイズが発生する。そのため従来は人工光源を撮影場所に設置することで光の不足を補っていた。本研究では、高速撮影問題をアンダーエクスポージャー画像のノイズ除去問題として捉え、深層学習によりアンダーエクスポージャーノイズを除去することで、短時間露光時の問題を解決する手法を提案している。一般的な低フレームレートのカメラ用の学習済みノイズ除去モデルはアンダーエクスポージャーノイズを想定したモデルではないため、それらを直接ハイスピードイメージングに適用してもアンダーエクスポージャーノイズを効果的に除去できない。よってアンダーエクスポージャー画像を用いたノイズ除去モデルを新たに学習する必要があるが、そのためにはハイスピードカメラでさまざまな環境下で撮影した大量の画像が学習データとして必要である。このデータ収集の負荷を軽減するために、本研究では			

氏名	庄 緯豪
----	------

一般的な低フレームレートカメラで撮影された大量の画像データをもとに、アンダーエクスポート画像をシミュレーションする手法を提案している。提案手法ではアンダーエクスポートノイズを被写体依存のノイズ成分とハイスピードカメラ依存のノイズ成分に分けて考え、後者のみを実際にハイスピードカメラを用いて収集し、前者は通常のカメラで撮影された画像から既存の信号処理技術によってシミュレーションすることで、深層学習モデルの学習時に十分な量のアンダーエクスポート画像の生成を実現している。実験の結果、提案手法は人工光源を設置せずとも、既存の方法よりも大幅に速いシャッタースピードで明瞭度の高い画像を記録することが可能であることを示した。定量評価結果として、提案手法は従来手法と同じ明瞭度を保ったまま、約 100 倍のシャッタースピードを実現できることを示している。

第三章では、動画再生時にフレーム間の情報を補間することで、記録するフレーム数を減らしても高フレームレートで再生可能にする手法を提案している。従来のビデオフレーム補間手法は、オプティカルフローに基づいて、観測された複数フレームのピクセル強度から中間フレームのピクセル強度を推定している。しかし、従来の方法は遮蔽や外部照明の変化などにより誤りを起こしやすいという問題がある。本研究では、動画データの時系列、および画像座標の 3 変数を入力すると、それに対応する画素値が出力されるニューラルネットワークを構築する手法を提案している。一般的な深層学習アプローチは事前に大量のデータを用いて学習させるのに対して、提案手法は事前学習を一切行わずに、フレーム補間を行うターゲットの動画データのみを用いてニューラルネットワークの最適化と補間処理を行うことが特徴である。提案手法では、ターゲットの動画データのみを用いて、時刻および座標とその画素値との対応をニューラルネットワークにより学習する。そして学習されたニューラルネットワークに対して、補間したい時刻および座標を入力すると、補間された画素の情報が出力されるという仕組みである。これは、観測済みの時系列データを何らかの関数で近似させる処理と類似の処理であり、近似関数にニューラルネットワークを使ったものとしてみなすことができる。実験において、複数のデータセットをベンチマークとして使用し、既存の state-of-the-art のモデルと比較した結果、優位な性能が得られることを示している。

第四章では三つ目の提案手法として、メモリ消費の削減手法について述べている。二つ目の提案手法によりフレーム補間が可能になれば、記録するフレーム数を削減することができるため、(空いたストレージに) 解像度の高い画像を記録することが可能となる。しかし深層学習ベースの画像処理システムでは高解像度画像の処理に大量のメモリを消費するという課題がある。そのため本研究では、畳み込みニューラルネットワーク (CNN) の推論段階でメモリ消費を削減する方法を提案している。提案手法は、高解像度画像を複数のタイルに分割し、分割したタイルを独立に CNN で処理し、処理結果を結合することでメモリの消費を抑えている。ただし分割しない本来の CNN の計算結果と提案手法による計算結果が一致することを保証するためには、タイル間でオーバーラップを持たせることが必要であり、これが処理時間のオーバーヘッドとなる。本研究では、CNN 内のどの層までをタイル化すると、必要なオーバーラップがどれだけ増え、それにより処理時間のオーバーヘッドがどれだけ増加するかを理論的に示し、タイル化を行う最適な層を実験的に求めている。実験では、PyTorch の実装において、MobileNetV2 や ResNet50 などの代表的な CNN モデルに対して、大幅なメモリ消費削減の実現性を示している。

第五章では、本論文の結論および今後の課題について述べている。

以上のように、本研究は制約された環境および計算条件におけるハイスピードイメージングシステムについて、その 3 つの課題：不十分な光源により発生するノイズの除去、フレーム補間、メモリ消費の削減について研究したものであり、ハイスピードイメージングのための深層学習アプローチについて重要な知見を得たものとして価値ある集積である。提出された論文はシステム情報学研究科学学位論文評価基準を満たしており、学位申請者の庄 緯豪は、博士 (工学) の学位を得る資格があると認める。