Physics-Informed Optical Flow Estimation via Algorithm Unrolling

Brad K. Rice
Advanced Multimedia Research Laboratory
University of Wollongong
Wollongong, Australia

Philip O. Ogunbona
Advanced Multimedia Research Laboratory
University of Wollongong
Wollongong, Australia

***Abstract*— The abstract gives away the results and lets the audience decide if they wish to read the document to understand how the author has reached those results. This means the abstract needs to be informative and descriptive.**

Roughly the structure should be:

- **25% setting the context and problem (2-3 sentences).**
- **Method (25%)**
- **The principle finding, its novelty and implications (50%, approximately 5-6 sentences).**

Update these index terms to appropriate terms

***Index Terms*—Physics-Informed Neural Network, Optical Flow, PINN, Deep Learning, Computer Vision, Algorithm Unrolling.**

I. Introduction

OPTICAL FLOW refers to the pattern of apparent motion of objects in a sequence of images. Optical flow describes how pixels in an image move between consecutive frames, allowing for an understanding of the dynamics and motion of objects, the camera or both. Optical flow is a foundational concept within computer vison and is core for a wide variety of applications including; stereo imaging, depth map estimation, 3D scene reconstruction, frame interpolation, scene understanding, applications to autonomous vehicles and other machine vision applications.

Importantly, optical flow allows for a sense of perceptual understanding and cognitive mapping through correlating visual flow to the motion of objects that are captured within a video feed. By accurately estimating optical flow, object movements can be tracked, changes can be detected, and an analysis of the flow of visual information can be conducted. Various applications can be enhanced with the additional spatiotemporal information, especially from highly dynamic scenes. A rudimentary example; estimations of the velocity and trajectory of an object of interest can be made allowing for decisions to be made by an autonomous system to avoid collisions.

There are two divergent approaches to estimating dense optical flow. Traditional methods, which include carefully handcrafted partial differential equations that are posed as an optimisation problem to solve. This leverages the framework of variational calculus where the phenomena is formulated as an energy functional to minimize. And deep learning approaches through deep neural networks used to directly regress the optical flow field.

Convolutional Neural Networks (CNNs) have revolutionised the deep learning approach to optical flow estimation, and with it's success, shifted research focus from traditional approaches to deep learning. Errors within the estimations of the optical flow field appear to diffuse throughout the networks architecture as opposed to errors accumilating for traditional methods [?], [?]. The current, state-of-the-art techniques, such as RAFT [?], heavily rely on CNNs, and nearly all leading methods integrate deep learning architectures into their frameworks. This paradigm shift has resulted in significant advancements in optical flow accuracy and performance, showcasing the effectiveness of deep learning's ability to capturing intricate motion patterns and improving flow estimation capabilities.

However, deep learning methodologies often require copious amounts of data to learn from. Collecting and annotating the large pools of data that is required to train these architectures is expensive and remains a major hurdle. Many of the benchmark datasets such as the KITTI [?], MPI-Sintel [?] and Middlebury [?] collections have limited examples for robust deep learning models to be trained for real world applications. Due to the difficulty in annotating ground truth information, many of these benchmarking datasets are synthetically generated where significant gaps in performance between real world datasets exists.

The general trend for improving deep learning methods appears to be building larger and more complex architectures or by training on more data. If this trend continues, progress will be limited to only those with sufficent computational resources. Instead, this can be mitigated by intelligently exploiting the rules of the system, *the physics*, and informing the deep learning architecture of this information. The conjecture is that Physics-Informed Neural Networks (PINNs) should provide substantial improvement to performance in a small data paradigm. PINNs can provide an interesting and attractive alternative to large scaled models trained on large scaled datasets. Providing justification to this idea, Raissi and Karniadakis have been successful in showcasing the power PINNs have as forward and inverse solvers of physics-based problems,

B. K. Rice is with the School of Computing and Information Technology, University of Wollongong, Australia.

P. O. Ogunbona is with the School of Computing and Information Technology, University of Wollongong, Australia.

Manuscript received (put the relevant date here date); revised July 14, 2024.

in a wide variety of areas, with strong generalisation. They demonstrate this where the physics is partially understood and with *only* small amounts of data to support the *known* physics [**?**], [**?**], [**?**], [**?**].

Algorithm unrolling, or unfolding, is a method where iterative algorithms are transformed into a chained series of the algorithm's primitive iterations. Each of the iterations becomes forward pass of a hidden layer within a neural network. Any of the parameters that are manually set in the traditional algorithm can be estimated throughout the unrolled variant, providing additional efficiency and flexibility. As a consequence, the unrolled layers in a neural network become intepretable and create a strong inductive bias. LeCun and Gregor introduced the Learned Iterative Shrinkage-Thresholding Algorithm (LISTA) as an unrolled version of ISTA [**?**] obtaining better performance than the traditional ISTA with fewer iterations, opening up an end to end learning framework for classical iterative techniques. Subsequently, algorithm unrolling has increased in popularity, shown significant advances and strong performance in solving inverse imaging problems by merging deep learning architectures with traditional optimisation approaches [**?**], [**?**], [**?**].

By unrolling algorithms that solve the physics of the system, an understanding of the governing rules can be directly embedded into the networks architecture. This allows the model to leverage the interpretability and structure of classical methods while enhancing them with the adaptability and data-driven insights of neural networks. This amalgamation not only improves the convergence and accuracy of the solutions but also provides a framework for integrating domain knowledge directly into neural networks, leading to more robust and generalizable models.

In this work, we propose a novel physics-informed architecture that directly embeds knowledge of optical flow by unrolling a gradient descent optimization designed to solve optical flow. The integration of optical flow knowledge into the proposed deep learning architecture not only enhances generalization, particularly within a small data paradigm, but also facilitates elucidation of the model's estimation of flow patterns. modify this statement as we confirm/reject/ this hypothesis (or reveal another finding etc.) .

In the following section, section **??** we provide preliminaries by presenting the theoretical foundations and mathematical tools that underpin optical flow. Section **??** presents the related works. Sections **??** and **??** show the methodology and the experiment design. Sections **??** and **??** present the results and provide a discussion. Sections **??** and **??** conclude and mention future directions for research to persue.

II. Preliminaries

A. Optical Flow Derivation

The intensity of a pixel, I, at position (x, y, t) can be represented by $I(x, y, t)$. At a small step in time, δt, the pixel will have moved by δx and δy amounts respectively. Assuming that the brightness is consistent between frames, the intensity of the new frame can then be represented as $I(x+\delta x, y+\delta y, t+\delta t)$.

By means of Taylor expansion, assuming small displacements, we can obtain;

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t) \\ + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \mathcal{O}(I^2)$$

By truncating the higher order terms it follows that;

$$\begin{aligned} 0 &= \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t \\ 0 &= \frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t}\frac{\delta t}{\delta t} \\ 0 &= \frac{\partial I}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial I}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial I}{\partial t} \\ 0 &= I_x u + I_y v + I_t, \end{aligned}$$

resulting in the *linearised* Optical Flow equation;

$$0 = \nabla I \cdot \vec{V} + I_t. \quad (1)$$

Equation **??** contains two unknowns for one known and therefore there exists an infinite number of solutions. (I don't like this sentence, but I'm unsure how to change it to convey the same message) This is known as the *aperture problem*. This aperture problem arises from the inherent ambiguity in determining the true motion direction of objects when only partial or limited information is available, i.e. the size of the aperture is too small. To compute the *true* optical flow, additional information is required, normally by introducing additional constraints.

B. Variational Methods

Among the most successful and widely used traditional techniques for optical flow estimation are variational methods. These schemes estimate optical flow by minimizing an energy functional. By formulating the problem as an energy minimization task, variational techniques leverage the mathematical framework of variational calculus to optimise the flow field and capture complex motion patterns. In particular, variational techniques are attempting to minimize the energy functional F such that we obtain;

$$\inf \left\{ F[u] = \int_{\Omega} f(x, u(x), \nabla u(x))\, \mathrm{d}x \right\}. \quad (2)$$

This typically found through solving the Euler-Lagrange equation, as shown in equation **??**, either directly or through numerical methods, which are predominantly iterative in nature. (COMMENT TO PHILIP: more opportunities for us to unroll other variational methods? Or even a generic framework for solving PDEs through unrolled iterative techniques?) What about an unrolled RK4 for ODEs or a generic finite element solver for PDEs? Would it be possible to do this irrespective of the designed mesh (i.e. infinite resolution)? The more that I'm diving into unrolling, I'm thinking that a key concept for my overall thesis lies in unrolling and numerical solutions to

PDEs (at least the key connections here from an analysis point of view) with the application to computer vision.

$$\frac{\partial f}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial f}{\partial \dot{u}} = 0. \tag{3}$$

The choice of f has the effect of determining what image problem we are attempting to solve. The Horn-Schunck method is one of the foundational variational methods that enforces a global constraint of smoothness, in the flow sense, to solve the aperture problem. The method constructs the energy functional with regularization parameter α to be minimised;

$$F[u] = \iint (I_x u + I_y v + I_t)^2 + \alpha(||\nabla u||^2 + ||\nabla v||^2)\ \mathrm{d}x\mathrm{d}y.$$

This minimisation is reduced through the Euler-Lagrange equations to form the system of PDEs to solve over;

$$0 = \nabla I(\nabla I \cdot \vec{V} + I_t) - \alpha^2 \Delta \vec{V}. \tag{4}$$

Most variational techniques, such as the Horn-Schunck presented above, have the underlying weakness from their assumptions of brightness constancy and smoothness that don't hold true in practice. Considering the limitations of traditional variational techniques in handling real-world challenges, alternative approaches leveraging deep learning have gained attention for optical flow estimation.

C. Lucas-Kanade Optical Flow Estimation

Lucas and Kanade approached the aperture problem through the assumption that within a local neighbourhood of pixels, flow should be constant. The key idea behind this method is to construct a set of linear equations from a window to solve via the least squares method. Using this additional pixel information, selected from an $m \times n$ image window, the method forms the system of equations;

$$\begin{aligned} 0 &= I_x(p_1)u + I_y(p_1)v + I_t(p_1) \\ &\vdots \\ 0 &= I_x(p_{mn})u + I_y(p_{mn})v + I_t(p_{mn}), \end{aligned}$$

equivalently;

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ \vdots & \vdots \\ I_x(p_{mn}) & I_(p_{mn}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(p_1) \\ \vdots \\ I_t(p_{mn}) \end{bmatrix},$$

which we will notate using the boldface font to mean the system;

$$\nabla \mathbf{I} \cdot \vec{V} = -\mathbf{I}_t. \tag{5}$$

The solution to this systems comes though solving the least squares problem;

$$\begin{aligned} 0 &= \nabla \mathbf{I} \cdot \vec{V} + \mathbf{I}_t \\ -\mathbf{I}_t &= \nabla \mathbf{I} \cdot \vec{V} \\ -(\nabla \mathbf{I})^T I_t &= (\nabla \mathbf{I})^T \nabla \mathbf{I} \cdot \vec{V} \\ -\nabla \mathbf{I}^T \mathbf{I}_t &= (\nabla \mathbf{I}^T \nabla \mathbf{I}) \cdot \vec{V} \\ -(\nabla \mathbf{I}^T \nabla \mathbf{I})^{-1} \nabla \mathbf{I}^T \mathbf{I}_t &= (\nabla \mathbf{I}^T \nabla \mathbf{I})^{-1} (\nabla \mathbf{I}^T \nabla \mathbf{I}) \cdot \vec{V} \\ \vec{V} &= -(\nabla \mathbf{I}^T \nabla \mathbf{I})^{-1} \nabla \mathbf{I}^T \mathbf{I}_t. \end{aligned}$$

For a solution to exist, $\nabla \mathbf{I}^T \nabla \mathbf{I}$ is required to be invertiable, i.e. $\det(\nabla \mathbf{I}^T \nabla \mathbf{I}) \neq 0$. Additionally, the eigenvalues of $\nabla \mathbf{I}^T \nabla \mathbf{I}$, λ_1 and λ_2 to have sufficiently large magnitude to alleviate any stability issues. The ratio and magitude of λ_1 and λ_2 can be used to infer if the current pixel is a corner, edge or neither, where corners and edges give more reliable optical flow information. Fullfilment of these conditions is a key weakness of the least squares approach.

D. Algorithm Unrolling

Algorithm unrolling is technique that converts traditional iterative optimization algorithms into trainable deep neural networks. This process involves expressing each iteration of an algorithm as a layer within a neural network architecture. The unrolling technique begins with an optimization algorithm, that involves repeating a series of operations until convergence is achieved. By unrolling these iterations, each step of the algorithm is mapped to a corresponding layer in a neural network. This transformation allows the parameters of each layer, originally the fixed parameters of the iterative steps, to be learned during the training process. Unrolling permits the network to optimize these parameters adapting them for better performance on specific tasks. The general process is to take each of the update steps;

$$x^{(i+1)} \leftarrow \mathcal{F}(x^{(i)}, \Theta), \text{ for } i \in \{1, \ldots, n-1\},$$

and stack each iteration into n layers of a neural network.

(Should I have a little diagram showing this?)

A primary advantage lies in its ability to accelerate convergence. Traditional iterative methods may require a significant number of iterations to reach an optimal solution, especially in high-dimensional or complex problem spaces. When these iterations are unrolled into a finite number of neural network layers, the network is trained to optimise the parameters for each unrolled iteration, ensuring efficient convergence. Moreover, algorithm unrolling provides a structured method to incorporate domain-specific knowledge, *the physics of the system*, into the learning process as a strong prior. Each layer retains the interpretability of its corresponding iterative step, offering insights into the network's functionality and decision-making process, giving attributes of interpretability, explainability and ensuring the results conform to the known physics.

A classical and concrete example of algorithm unrolling can be made by observing the Iterative Shrinkage Thresholding Algorithm, ISTA. The classical ISTA aims to optimise the problem;

$$\min \frac{1}{2}||Ax - y||_{L_2}^2 + ||\lambda x||_{L_1}, \tag{6}$$

by iterating until there is convergence for a solution of x, using the update rule;

$$x^{(k+1)} \leftarrow S_\alpha\Big(x^{(k)} - \frac{1}{L}A^T(Ax^{(k)} - y)\Big), \tag{7}$$

where $S_\alpha(\star) \triangleq \mathrm{sign}(\star) \cdot \max\{|\star| - \alpha, 0\}$, is the shrinkage operator, (soft-thresholding operator with threshold α). Instead the unrolled version can be designed by stacking K iterations together, forming a K-layer deep network. Implicit substitutions are made, $A_t = I - \frac{1}{L}A^T A$ and $A_e = \frac{1}{L}A^T$, where I is the identity matrix, creating the parameters A_t, A_e and α for the network. Training is performed, through a loss function ℓ, defined as;

$$\ell(A_t, A_e, \alpha) = \frac{1}{n}\sum_{i=1}^{n} \left|\left|x^{(i)} - \hat{x}^{(i)}(y^{(i)}; A_t, A_e, \alpha)\right|\right|_2^2, \quad (8)$$

by forming a sequence of vectors $y^1, \dots, y^n$ and their relative solution vectors $x^1, \dots, x^n$, as the ground truth.

E. Physics Informed Neural Networks

Physics Informed Neural Networks (PINNs) are a novel class of neural network that merge deep learning techniques with the laws of physics to solve complex physical problems. By incorporating known physics as a prior for the network, PINNs offer a framework for efficient and precise modelling, simulation, and prediction in diverse scientific domains. This integration of domain knowledge enhances the interpretability, generalisability, and robustness of neural network models. Their strength lies where small amounts of data is available and there is a partial understanding of the physics involved.

Neural network architectures can incorporate physics through three main methodologies. Firstly, the direct embedding of mathematics within the network's architecture, known as inductive bias, enables the network to incorporate prior knowledge of the system. This can be done via embedding phyical laws such as the conservation of energy, or by embedding iterative solvers designed to solve differential equations that explain the system. The method presented in this work induces bias into the networks architecture by unrolling a gradient descent method to solve the physics of optical flow. Secondly, by providing the network with physics-driven information, such as leveraging observational biases, guides the network towards appropriate inference, such as exploiting symmetries in molecular structures or in imaging sequences. Lastly, a learning bias can be introduced by penalizing the network when its learning deviates from the underlying physical laws. This approach often involves incorporating the mathematical model underlying the system directly into the network's loss function, causing the network to learn by optimising over the physics.

NOTE: Two things;

1) Do I need these preliminaries for this audience (I feel as if these are too basic and are only here for completeness)? Would it be weird to only have physics-informed neural networks and algorithm unrolling without the two methods?
2) I have included Horn-Schunck and Lucus-Kanade because the plan is to unroll both and to compare to the traditional.

III. Related Work

Optical flow estimation has been a long-standing research topic in computer vision, with various techniques contributing to its advancement. Due to it's foundational nature and importance, it is still of contemporary research interest [?], [?], [?], [?]. Currently, there are two main approaches to estimate optical flow. Traditional techniques that typically leverage the calculus of variations, minimising a handcrafted energy functional. One of these foundational techniques, the Horn-Schunck method [?], has been widely used for estimating optical flow by assuming brightness constancy and smoothness constraints to solve the aperture problem. More recently, deep learning-based approaches, such as FlowNet and its variants [?], [?], [?], [?], have become the norm by leveraging CNNs to learn complex motion patterns directly from large-scale training data, leading to significant improvements in accuracy and robustness of optical flow estimation.

Both variational and deep learning approaches suffer from distinct disadvantages. A major challenge for variational techniques is the assumption of brightness constancy, which assumes that the brightness of pixels remains constant between consecutive frames. In practice, variations in lighting conditions, occlusions, and motion blur can violate this assumption, leading to inaccuracies in flow estimation. Another challenge is the smoothness constraint, which assumes that neighbouring pixels have similar motion. This assumption may not hold in the presence of discontinuities, large displacements or complex motion patterns, leading to incorrect flow estimates. Motion discontinuities pose a challenge to dense optical flow algorithms. In variational optical flow methods, the excessive smoothing of flow discontinuities contributes significantly to the overall error [?]. The complexity of the energy functional needed to explain the variety of motion patterns to provide a generalisable solution inhibits application compared to deep learning techniques for accurate flow estimation (is this statement too strong to mention without a reference?).

Deep learning methods have been found to perform particularly well for large displacements and complex motion patterns, however, issues still exist [?]. Due to the size and complex architecture of deep learning architectures, they have a large number of parameters to learn causing a large memory footprint and require copious amounts of well labelled data learn from. The current datasets used for benchmarking; KITTI, MPI-Sintel and Middlebury do not have sufficient examples for a deep learning model to be robust for real world applications. Methods of generating synthetic optical flow examples have been used to bridge this gap, however, these data samples struggles to represent issues found in practice, such as motion blur and lighting variations [?].

Physics informed learning methods have emerged as a promising approach in a wide variety of applications by integrating the underlying physical principles into the learning process. The broader research community has been lauding over the introduction of physics informed machine learning (PIML) methodologies since they were introduced by Raissi [?]. Within their seminal paper, hidden fluid flows were learned by a PINN showcasing their power to generalise well in the presence of noisy and low fidelity information [?]. This spawned multiple branches for PIML to explore, most of which have found PIML methods have increased the powers of deep learning methods, especially in a small data paradigm [?].

Recently the term physics-informed computer vision (PICV) has been coined to describe the emerging domain of applying physics-informed models with computer vision tasks [?]. Physics-informed computer vision models exploit the benefits of having information gained from the underlying physics of the system and data to validate it's inference. The PICV domain is still in it's infancy and has many opportunities for rapid advancement. Examples of recent PICV applications include; superesolution [?], image generation [?], image reconstruction [?], image segmentation [?], [?] and crowd analysis [?]. One such notable application includes inferring fluid flows from 4D-MRI imaging for the purposes of understanding arterial wall stress leveraging the well known Navier Stokes equations for incompressible fluid flows [?]. Estimations of the fluid flow vector field allows for the pressure and velocities to be estimated for advanced non-invasive diagnositcs for vulnerable patients. There is a strong link between [?] and learning of optical flow, however, it must be noted that as of yet we have not found any literature that directly learns an optical flow estimation using a PIML method. The links between physics-informed learning and algorithm unrolling are begining to strengthen, and our work aids to justify the link between the two. With optimisation proceedures directly induced into the deep learning network, the argument is that the architecture can resolve the physics of the system and make effective use of this information to perform it's designed task.

Algorithm unrolling has emerged as part of the contemporary computer vision research community and has impacted the intersection of iterative optimization techniques and deep learning, increasing both efficiency and model performance. The concept, pioneered by Gregor and LeCun with the introduction of the Learned Iterative Shrinkage-Thresholding Algorithm (LISTA) [?], where this approach has been effectively employed in various domains, such as compressive sensing MRI [?], and continues to evolve. Widespread adoption is evidenced in a review by Monga et al. [?]. The review demonstrated the effectiveness that unrolled optimisation algorithms have in solving inverse problems and hold substantial benefits for model explainability. Additional benefits are realised by inducing prior knowledge into deep learning architectures, fullfilling the inductive bias definition of physics-informed learning, allowing for a principled approach to modelling the problem space with deep learning architectures that have prior understanding. Our work builds on these advancements by proposing a novel physics-informed algorithm unrolled deep learning architecture, specifically tailored for optical flow estimation. By unrolling gradient descent and proximal gradient descent, our method embeds knowledge of the physical constraints directly into the learning process, aiming to achieve enhanced generalization and accuracy even in data-scarce scenarios.

There is a strong belief we are the first to perform optical flow estimation using an unrolled gradient descent physics-informed, deep learning method. As such, this work contributes a novel approach by leveraging the physics of optical flow, and the flexibility and robustness of deep learning. By directly influencing the deep learning architecture by inducing optical flow knowledge, the aim is to enhance the current capabilities of optical flow estimation while opening up new avenues for research and applications into the nascent field of physics-informed computer vision.

IV. Methods

Here we should detail our methodology in detail.

Each image pair, and it's associated ground truth is split into '*patches*', to reduce the visual field of the architecture. (sliding window (guassian function) or apply the YOLO logic?). NOTE: we will need to define how we resolve the flow between patches? How do we reconstruct an image from a series of patches? (From the patches this means we could parallerise this process too). From these patches, for each pixel we will use a neighbourhood of surrounding pixels to form a linear system of optical flow equations to solve (Lucus-Kanade approach). This will be done via the unrolled proximal gradient descent physics-informed network.

Unrolled Proximal Gradient Descent

Suppose we have some noise in our measurements, i.e. those from an imperfect sensor or other sources of noise, modelled as $\varphi(\cdot)$. And let's reconsider the minimisation problem for our least squares problem (now in the L2 sense);

$$\min_{\vec{V} \in BV(\Omega)} \left|\left|\nabla \mathbf{I} \cdot \hat{\vec{V}} + \mathbf{I}_t\right|\right|^2_{L_2} + |\varphi(\cdot)|_{L_1}.$$

NOTE: $\varphi(\cdot)$ acts like a regulariser? If we have this in the L1 sense then this should promote sparsity right? Would this be better to be in the L2 sense, but we hope noise will be small so in the L2 sense it won't be effective at influencing the optimisation.

What if we unroll a proximal gradient descent for solving this least squares problem? Proximal gradient descent is used since I'm not convinced in reality we will have an everywhere differentiable function i.e. ($f \notin C_1(\Omega)$). Also consider the operator, $\text{prox}_\varphi(\star) \triangleq \arg\min_z || \star^{(i-1)} - z||^2_{L_2} + |\varphi(\cdot)|_{L_1}$. Update rules are;

$$\begin{aligned} z^{(i)} &\leftarrow \text{prox}_{\varphi}(\vec{V}^{(i-1)}) \\ \Rightarrow z^{(i)} &\leftarrow \arg\min_z ||\vec{V}^{(i-1)} - z||^2_{L_2} + |\varphi(\cdot)|_{L_1} \\ \vec{V}^{(i-1)} &\leftarrow z^{(i)} + \eta^{(i)} \nabla \mathbf{I}^T (\mathbf{I}_t - \nabla \mathbf{I} z^{(i)}). \end{aligned}$$

NOTE: we have a dynamic learning rate (stepsize), η_i. This can be learned at each step for an unrolled network.

The trick that we need to solve now is how do we find $\arg\min_z ||\vec{V}_{i-1} - z||^2_{L_2}$ efficiently? Is this where I could use a neural network to approximate?

A traditional loss function for optical flow could be described by the end point error (EPE).

$$\ell = \frac{1}{n}\sum_{i=1}^{n} |V_i - \hat{V}_i|_{L^2}$$

And the physics informed variant used to train our unrolled network is;

$$\ell = \frac{1}{n}\sum_{i=1}^{n} |V_i - \hat{V}_i|_{L_2}^2 + \frac{1}{N}\sum_{i=1}^{N} |\nabla I \cdot \hat{V}_i + I_t|_{L^2}^2$$

NOTE: removal of the EPE metric component of the loss function, and sole reliance on the physics loss could be used to train the model, and could be used to form a *self-supervised* approach to learning.

V. Experiment Design

Talk about the experiment design (if applicable)

VI. Results

Present the results that we obtain.

VII. Discussion

VIII. Conclusion

Discuss the consequences of the results, discussion surrounding the importance and whether it solves the problem appropriately or if there is still further work to be conducted.

IX. Future Work

Talk about what should be explored for the future.

Acknowledgment

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

X. Biography Section

Brad Rice is a doctoral candidate in the School of Computing and Information Technology at the University of Wollongong, Australia. His work focuses on '*Physics Inspired Deep Neural Networks for Computer Vision*'. Physics-informed learning has a broad applicability across various domains and allow for an amalgamation between first principles and machine learning methods. Brad is currently a student researcher in the Advanced Multimedia Research Lab, University of Wollongong and works in the energy and resources industry as a senior data scientist. Brad holds a Master of Computer Science (2021) and Bachelor of Mathematics (2019) from the University of Wollongong, where his research interests include computer vision, applied mathematical modelling, optimisation methodologies and machine learning.

Philip O. Ogunbona received the bachelor's degree in electronic and electrical engineering from the University of Ife, Nigeria, and the D.I.C. (electrical engineering) and Ph.D. degrees from Imperial College London. After his graduation, he has worked for many years at the University of Wollongong, Australia, and the Motorola Multimedia Center Laboratory. He is currently a Professor and the Co-Director of the Advanced Multimedia Research Lab, University of Wollongong, Australia. His research interests include computer vision, pattern recognition, medical imaging, and machine learning.