



HAL
open science

Can Simulated Retinal Wave pre-training improve Optical Flow Estimation?

Christos Kyriazis, Bruno Cessac, Hui-Yin Wu, Pierre Kornprobst

► **To cite this version:**

Christos Kyriazis, Bruno Cessac, Hui-Yin Wu, Pierre Kornprobst. Can Simulated Retinal Wave pre-training improve Optical Flow Estimation?. Neuromod meeting 2024, Jul 2024, Antibes, France. hal-04735162

HAL Id: hal-04735162

<https://inria.hal.science/hal-04735162v1>

Submitted on 14 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Summary

- **Goal:** To investigate how *simulated Retinal Waves (RW)* could improve machine learning models for **Optical Flow Estimation (OFE)**.
- **Why?** OFE models train primarily on synthetic data and not real world video, which can introduce biases. RWs are a fundamental step in the visual system development, especially in the motion-related circuitry.
- **Required Elements:**
 - RW biophysical **model** and **simulator**
 - Compatible Machine Learning **model** for OFE.

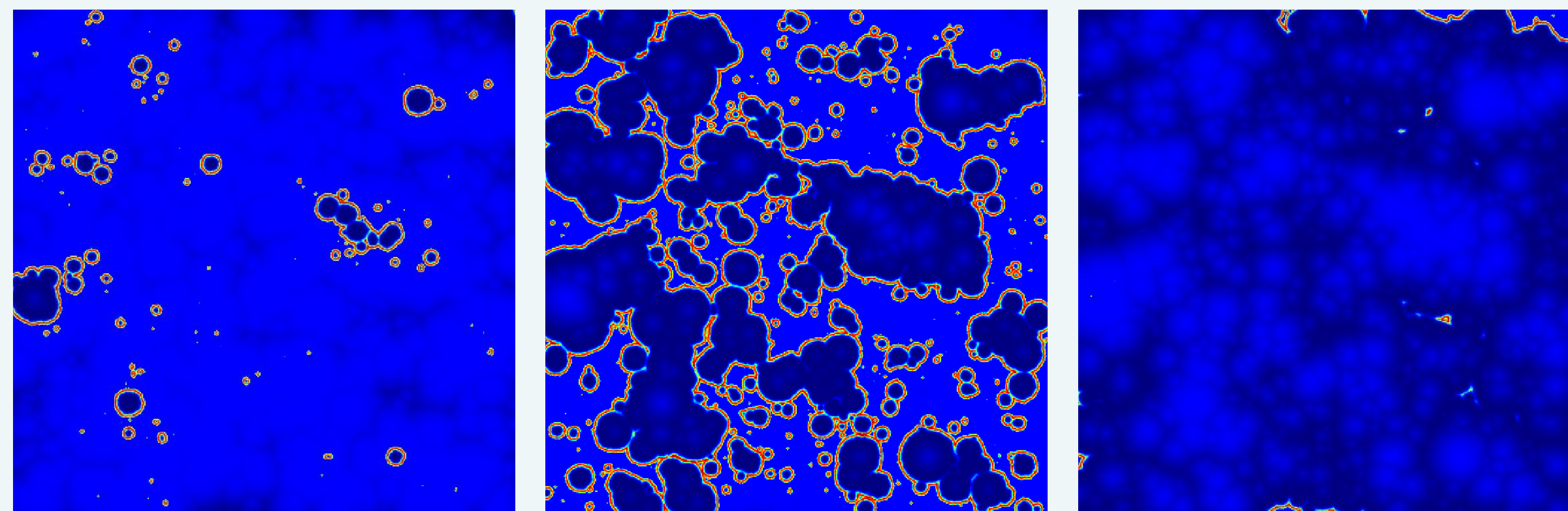


Figure 1. Simulated RWs visualised at 15sec increments in a 432x432 cell lattice during a 540sec long simulation. The SAC cells are colour-mapped by their intracellular calcium concentration; blue colour indicates a low concentration while red a high concentration.

1. Methods & Background

Modelling and Simulating Retinal Waves.

RWs are one of the early processes in visual system development, occurring around birth and playing a vital role in the **structuring of retinal circuits as well as their projections to the visual pathway**. These waves are spontaneous bursts of action potentials in the immature retina, forming activity waves in abstract patterns, and **prepare the visual system to detect motion before the eyes open**.

RWs occur in 3 stages and it is believed that the **stage II waves (cholinergic)** are the main contributor in synapse establishment among retinal ganglion cells and their projections to the brain. The neuron type that gives rise to these RWs are **Starburst Amacrine Cells (SACs)**, and are simulated using a biophysical model [1] in python using Brian2.

Optical Flow Estimation

OFE is a fundamental aspect of visual processing and plays a crucial role in various tasks such as object tracking and motion segmentation. Given two **consecutive images I_1 and I_2** , the optical flow is defined as the **displacement vector field that maps each pixel in I_1 to its corresponding pixel in I_2** .

Data-driven OFE models have been primarily using various synthetic datasets such as *FlyingChairs* [2] and *MPI-Sintel* [3]. The main reason for this is the difficulty of collecting densely labelled optical flow ground-truth, from real-world data.



Figure 2. A sample from the OFE dataset MPI-Sintel [3]. Taken from scene Bamboo-2; top: frames 34 and 35, left to right; bottom: corresponding optical flow visualisation. In the optical flow visualisation, colours indicate motion direction, while brightness represents motion magnitude.

2. Motivation

In recent years, **data-driven** OFE approaches have increasingly dominated benchmarks such as *MPI-Sintel*, relying heavily on large amounts of synthetic labelled datasets. We aim to explore the use of **RWs as easily accessible training data** for OFE, potentially improving:

- Training efficiency: Fewer iterations to reach a certain accuracy.
- Generalization efficiency: fewer data to reach a certain accuracy.

Artificial Neural Networks in Computer Vision learn differently than biological neurons during visual development. To our knowledge, no prior work has employed simulated RWs for motion-related machine learning tasks.

Can we use RWs for OFE? Not really...

Due to the lack of a mathematical formulation for ground-truth optical flow in the RW model, directly training OFE models using RWs is infeasible. Therefore, we chose a related motion task that captures temporal dynamics and spatial correlations: **Next Frame Prediction (NFP)** on RWs visualized with Calcium.

How to approach this idea? With Transfer Learning (TL).

In this approach, source and target **domains (D_S, D_T)** are RGB images while the source and target **tasks (T_S, T_T)** are different. T_S involves **predicting the next frame** given two consecutive images; while T_T involves **predicting the displacement** between two consecutive images.

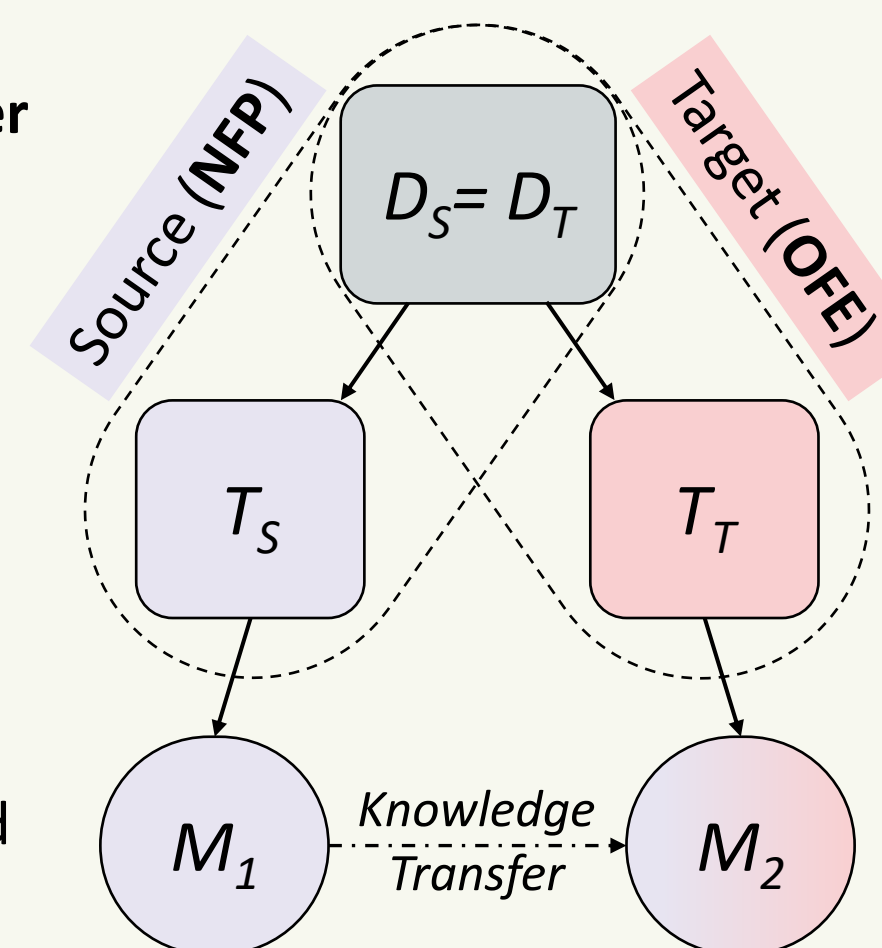
In TL, when $D_S = D_T$; $T_S \neq T_T$ we have **Task Transfer**

Definitions:

- Domain: $D = \{X, P(X)\}$
- Task: $T = \{Y, f(\cdot)\}$

where:

- X is the feature space
- X is the learning sample ($X \in \mathcal{X}$)
- Y is the label space
- $f(\cdot)$ is the predictive function which is learned from $\{x_i, y_i\}$ where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$



We want to observe **how** and if it is possible to **transfer the knowledge** between model M_1 , trained for **NFP**, and model M_2 , trained for **OFE**. This can be done in various ways, usually involving transferring subsections of M_1 to M_2 between the training stages.

4. Current Results

Step 3: NFP pre-training

Two models utilizing the two encoder variants of RAFT [4] are trained and tuned, on an initial dataset containing **simulated RW activity totaling ≈ 1 hour**.

Datasets of **different dt values** (frame intervals) are tested, along with standard machine learning hyperparameters.

The model is configured to optimize for **Mean Squared Error (MSE)** between predicted and ground truth I_2 of RW dataset sample.

For this step, we find no significant difference in MSE between 4 and 8 ConvGRU iterations ($< 0.5\%$ over baseline MSE).

Step 4: OFE training

To briefly test the effect of pre-training with this first model for NFP, **transfer learning** experiments are organized as follows:

- Feature Encoder weights and batchnorm weights are frozen
- Only Feature Encoder weights are frozen
- Only batchnorm weights are frozen

The encoder weights of different NFP models are compared, seen on Figure 4. Experiments are conducted on the NEF cluster, following the best training configuration for RAFT

As expected in this step, shuffling between samples (RW triplets) improves the NFP performance by $\approx 29\%$. This is due to increased data diversity and reduced temporal bias.

Figure 3 visualizes a **limiting factor** of NFP using RW: The model cannot predict the spontaneous start of RW clusters. We observe at the rightmost image sets that the level of error for these **unpredictable** RW segments, increases as the **frame interval (dt)** increases.

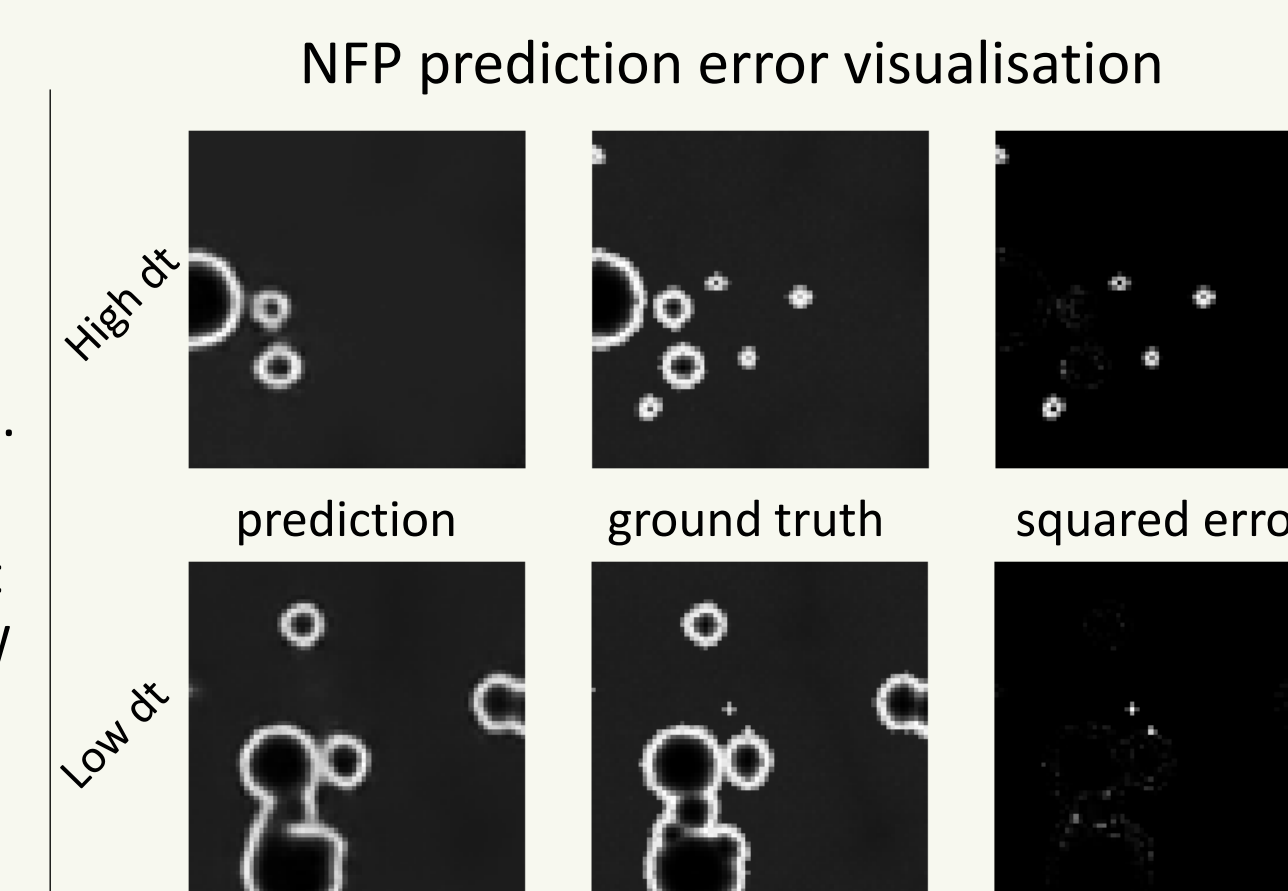


Figure 3. Images are 6x zoomed in

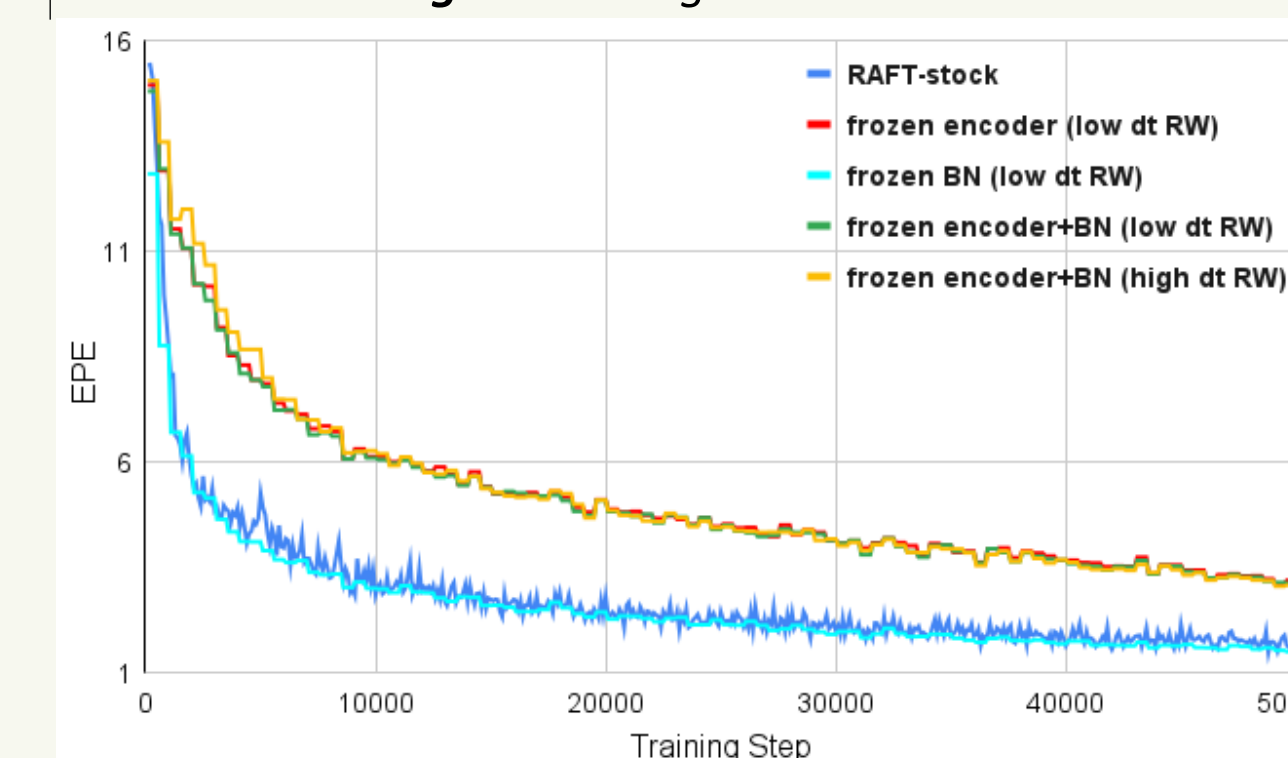


Figure 4. End-Point-Error during OFE training

3. Pipeline Overview

Step 1: RW dataset generation

An implementation of the SAC model was adapted and enhanced to generate an initial dataset at higher resolutions, benefiting from GPU acceleration. The chosen resolution (**432x432**) is relatively close to the FlyingChairs OFE dataset resolution. A range of model parameters is configured to obtain biologically-plausible simulations using 4-neighbor connectivity.

Step 2: Dataset pre-processing

The RW simulation data (Calcium) is normalized, ensuring consistent scaling across all values before organizing them into data frames. It is possible to accurately **filter the level of RW activity** sought-after, allowing to create dataset variations, with the same RW simulations.

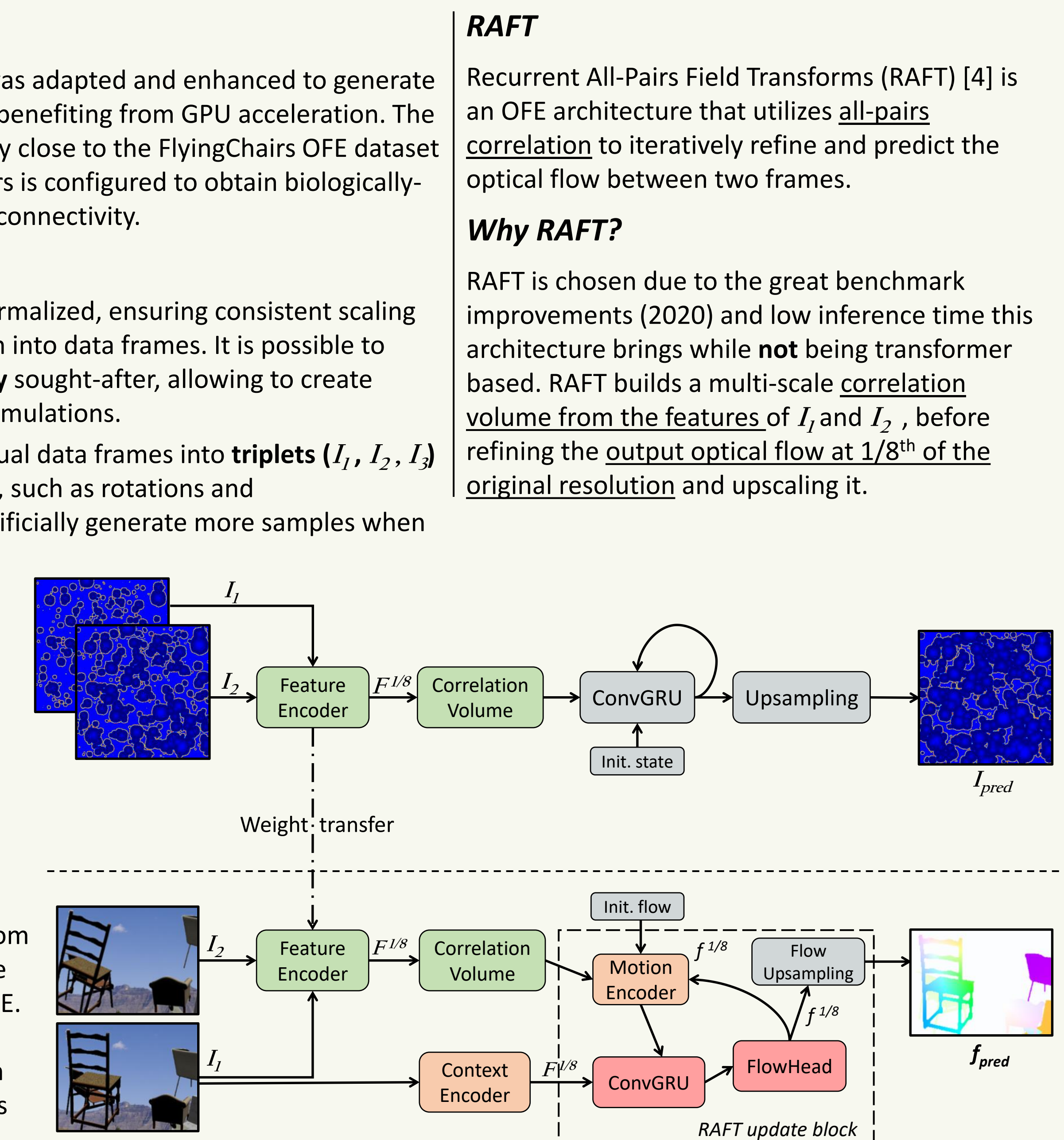
A PyTorch dataset organizes the individual data frames into **triplets (I_1, I_2, I_3)** for NFP. Data augmentation techniques, such as rotations and vertical/horizontal flips enable us to artificially generate more samples when needed.

Step 3: NFP pre-training

A NFP model is designed, utilizing components from RAFT annotated in **green**. The **Feature Encoder** learns to extract relevant features applied in correlation volumes, enabling the decoder to more accurately predict the next frame I_3 .

Step 4: OFE training

The relevant weights are transferred from the NFP model to stock RAFT during the **first training stage (FlyingChairs)** for OFE. There are several approaches to take during this step, using the weights from Step 3; this is a primary objective of this study. The hypothesis proposes that leveraging correspondence features between RWs can enhance OFE by providing RAFT with a beneficial initialization, potentially accelerating its generalization to other datasets.



Schematic diagram of the original RAFT architecture. Colours indicate **feasibility** and **ease of component reuse for NFP**; red denotes components where weight transfer is impractical; orange indicates significant modification required; green suggests compatibility with minimal modification

5. Interpretation and Next Steps

The motion features during NFP between two RW images do not bring any advancements with the currently employed transfer learning strategy. It can be further extrapolated that the training of the OFE model is **"slowed-down"**.

There can be several reasons for this, **to be investigated**:

- The source and target domains are not close enough, resulting in vastly different feature extraction from the encoder, during NFP and OFE.
- Not enough elements of RAFT are incorporated during the NFP pre-training step. RAFT uses an additional context encoder that could be used in step 3.
- The displacement in the RW data is too small.
- The motion variety present in the RW data is different than in OFE.
- Using triplets might negate wave temporal causality and ordering.

Having stated this, additional approaches should be explored to effectively utilize the motion found in simulated RWs within the RAFT framework.

References

- [1] B. Cessac and D. Matzakos-Karvouniari. *The non linear dynamics of retinal waves*. Physica D: Nonlinear Phenomena (2022).
- [2] A. Dosovitskiy et. al. *FlowNet: Learning Optical Flow with Convolutional Networks*. ICCV (2015)
- [3] D. J. Butler, J. Wulff, G.B. Stanley, M.J. Black. *A naturalistic open source movie for optical flow evaluation*. ECCV (2012)
- [4] Z. Teed and J. Deng. *RAFT: Recurrent All Pairs Field Transforms for Optical Flow*. ECCV (2020)