From Perceptual Filling-In to Stable Active Vision

Advanced Seminar in Computer Science

Supervised by Dr. Hadar Cohen-Duwek

September 2025

Efrat Friedrich

האוניברסיטה
הפתוחה

Introduction Outline

The Discrepancy of Visual Perception

Core visual phenomena

Biologically Inspired Systems

Research Progression Map

The Discrepancy of Visual Perception

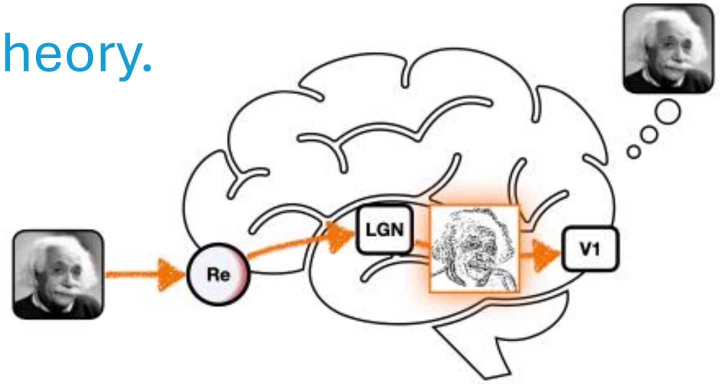
- The research addresses:
  1. The discrepancy between:
     - Subjective human experience: a complete, colorful, stable view.
     - Actual retinal input : incomplete, low-resolution, and highly dynamic due to constant eye movements (saccades).
  2. Limitations of Retinal Input:
     - The peripheral retina provides low resolution and weak color cues. It primarily transmits edge-related information rather than full surface details, stemming from the non-even distribution of photoreceptors.

The research aims to develop biologically plausible models that leverage neuromorphic systems (SNNs, event cameras) to model how the brain bridges this gap.

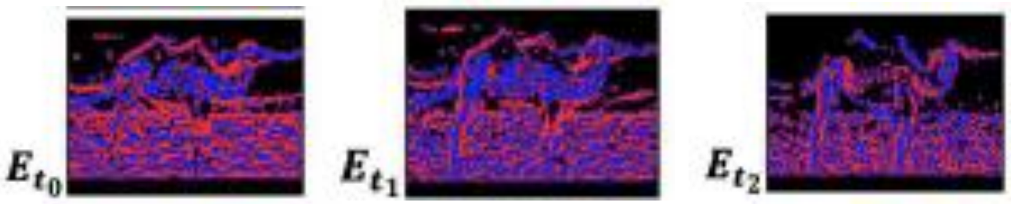
Core visual phenomena

1. Perceptual Filling-In

Isomorphic Theory.

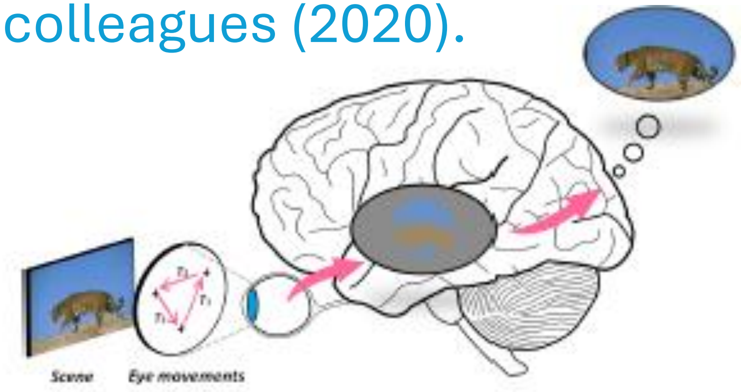


2. Efficient Image Reconstruction (events frames)



3. Peripheral Vision / Perceptual Colorization

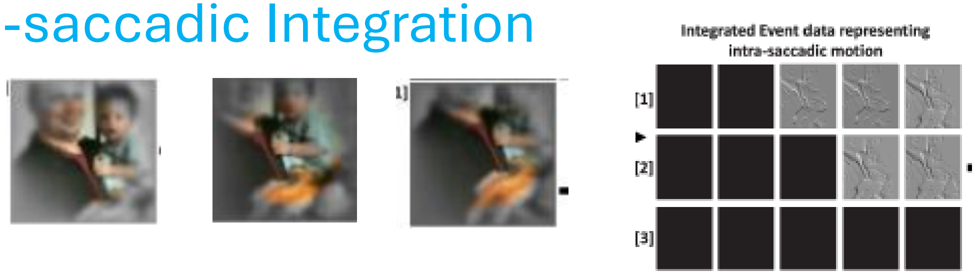
Cohen and colleagues (2020).



4. Saccadic Vision/ Visually Stable Perception

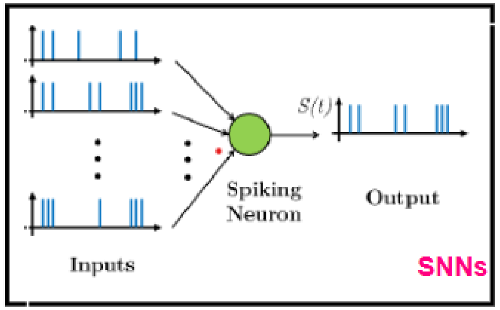
Corollary Discharge (CD) signals

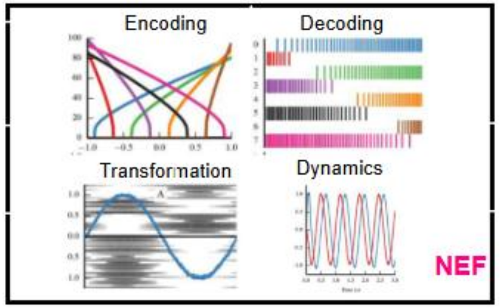
Trans-saccadic Integration

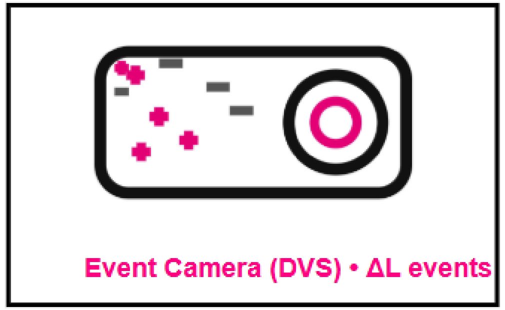


→ **The research computationally models these 4 biological visual phenomena**

Biologically Inspired Systems





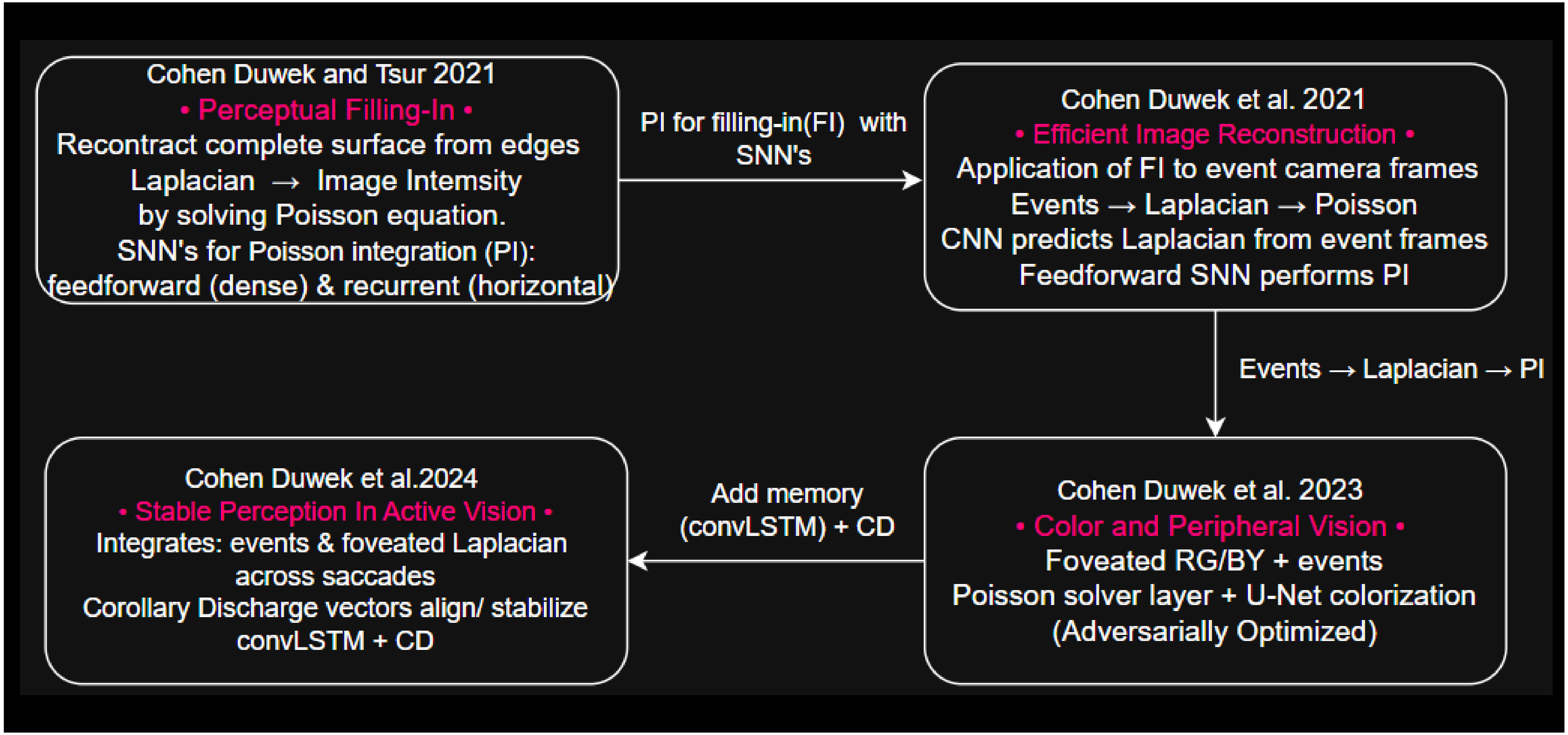


Event
Camera(AER)
↓
SNN (+ NEF)
↓
Output
(Reconstruction)

Together, these systems facilitate the development of **realistic and dynamic brain-inspired models** for core visual phenomena.

Research Progression Map

From Perceptual Filling-In to Stable Active Vision



Brief Overview of Each Paper

Foundational Work : Perceptual Filling-In Mechanisms [Source 1]
Application to Event Cameras: [Source 2]
Expanding to Color and Peripheral Vision: [Source 3]
Achieving Visual Stability in Active Vision [Source 4]
Overall Conclusion and Future Directions

1.Foundational Work :
Perceptual Filling-In Mechanisms

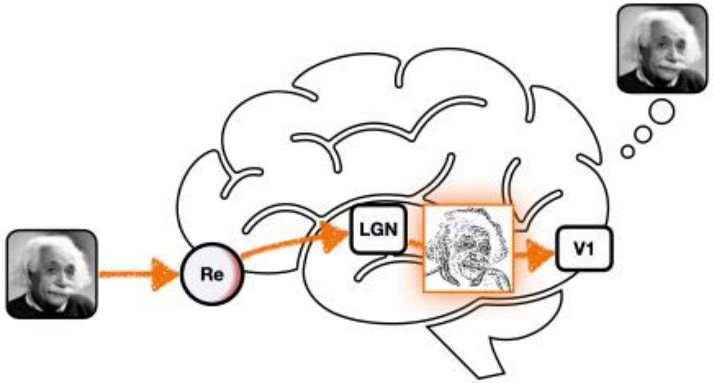
Biologically Plausible Spiking Neural Networks for Perceptual Filling-In

Cohen Duwek and Tsur (2021) [Source 1]

Focus: SNN for Poisson Integration , edges → surfaces

Poisson Equation $\nabla^2 I = -f$

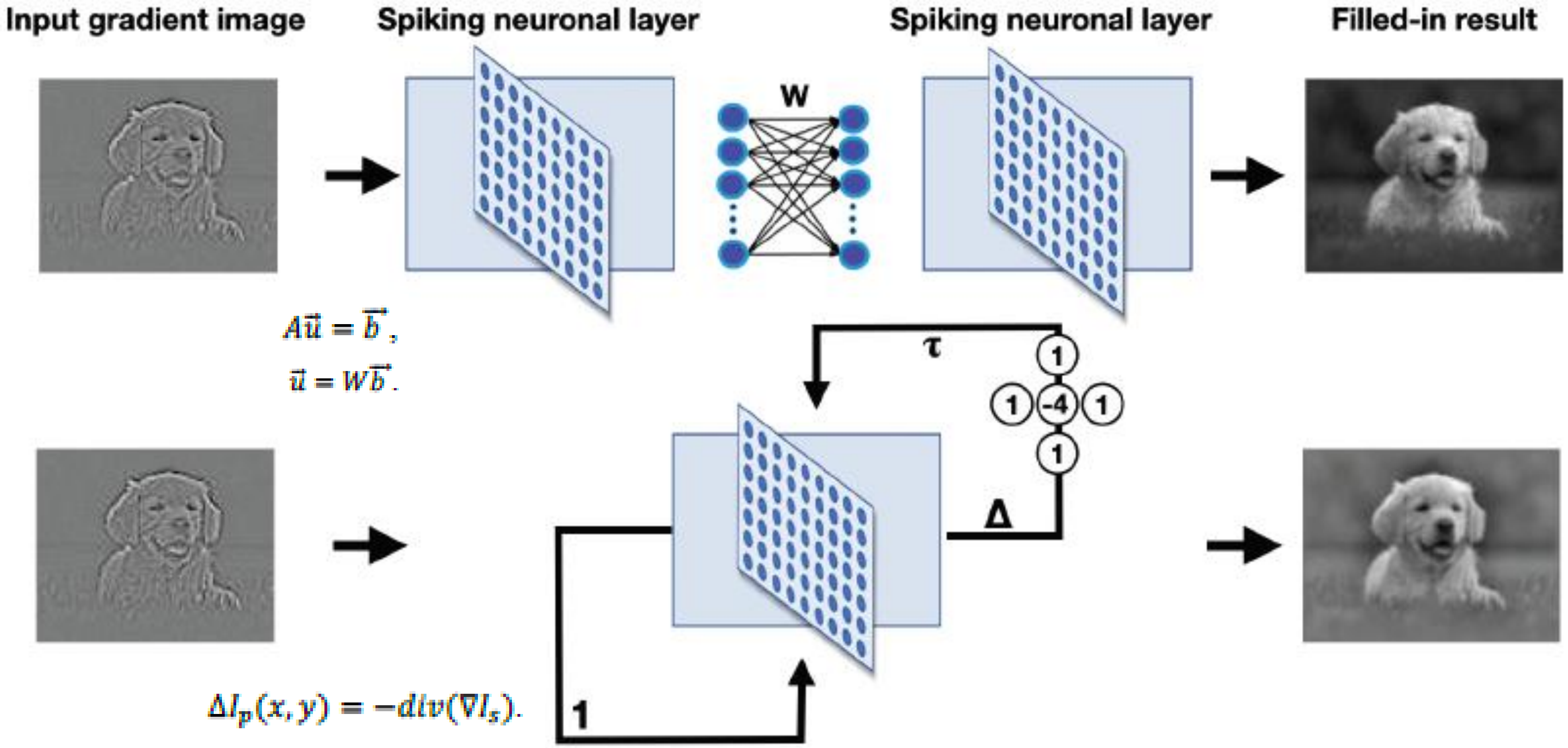
משוואה דיפרנציאלית חלקית



Model Architecture

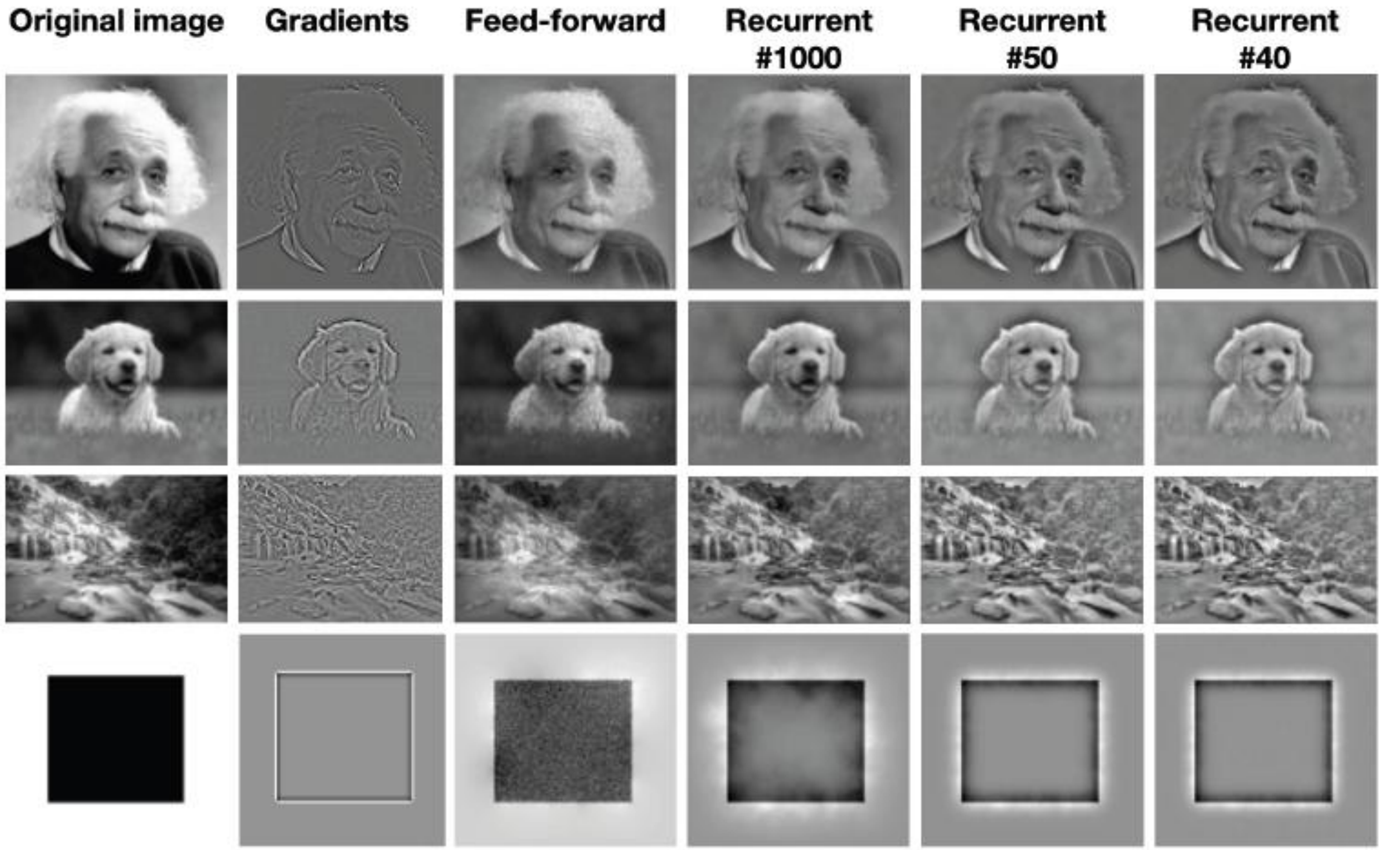
Laplacian → Poisson Integration (PI) → Image Intensity

- **Input** is the image Laplacian, representing retinal ganglion cell receptive fields
- **Feedforward SNN (top)** Dense connections of two spiking neuronal layers
- **Recurrent SNN (bottom)** Image is reconstructed iteratively over time through recurrent (horizontal) connections.

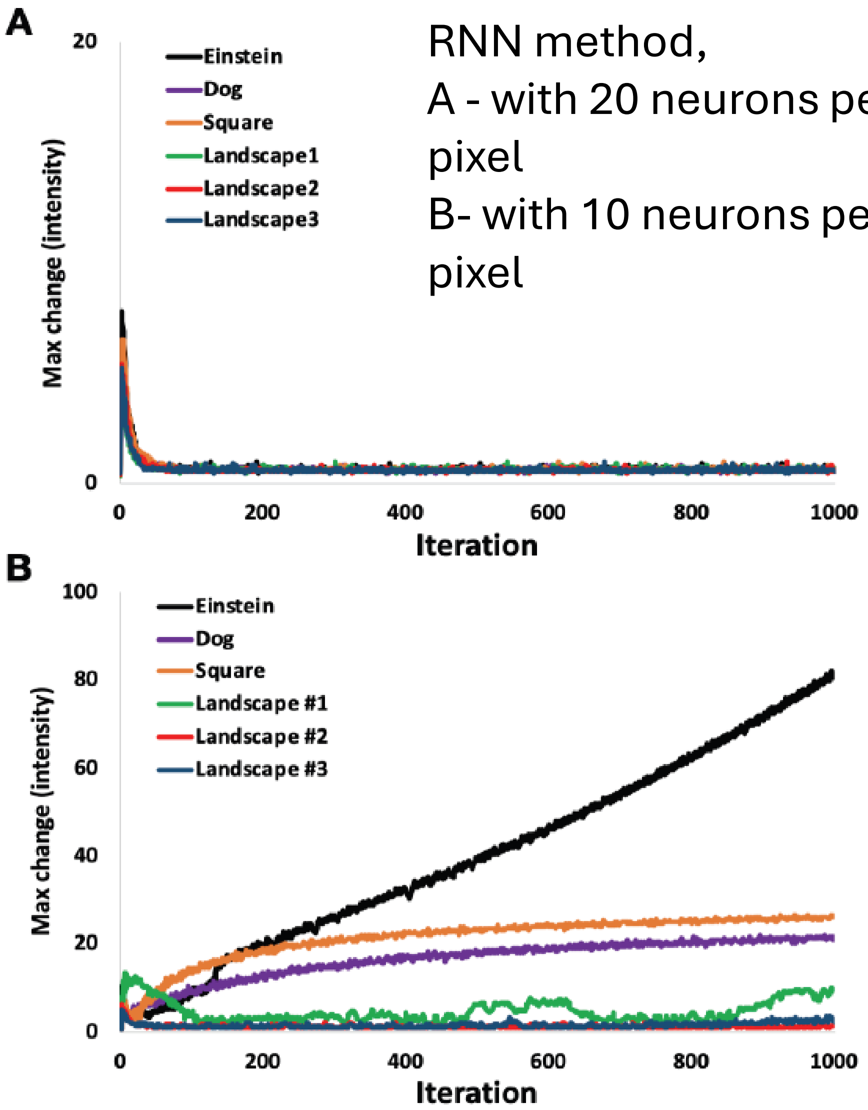


[Source 1]

Main Results



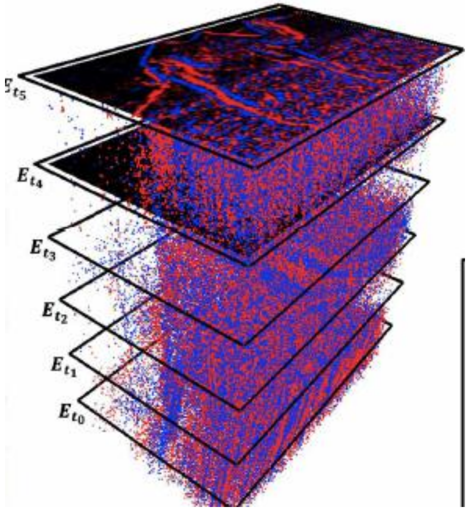
(Left) visually comparing the reconstructed images with the original images



RNN method,
A - with 20 neurons per pixel
B- with 10 neurons per pixel

y-axis absolute maximal change across all pixels. Convergence when it reduced to 0

[Source 1]



2. Application of perceptual filling-in to Event Cameras

Image Reconstruction from Neuromorphic Event Cameras using Laplacian-Prediction and Poisson Integration

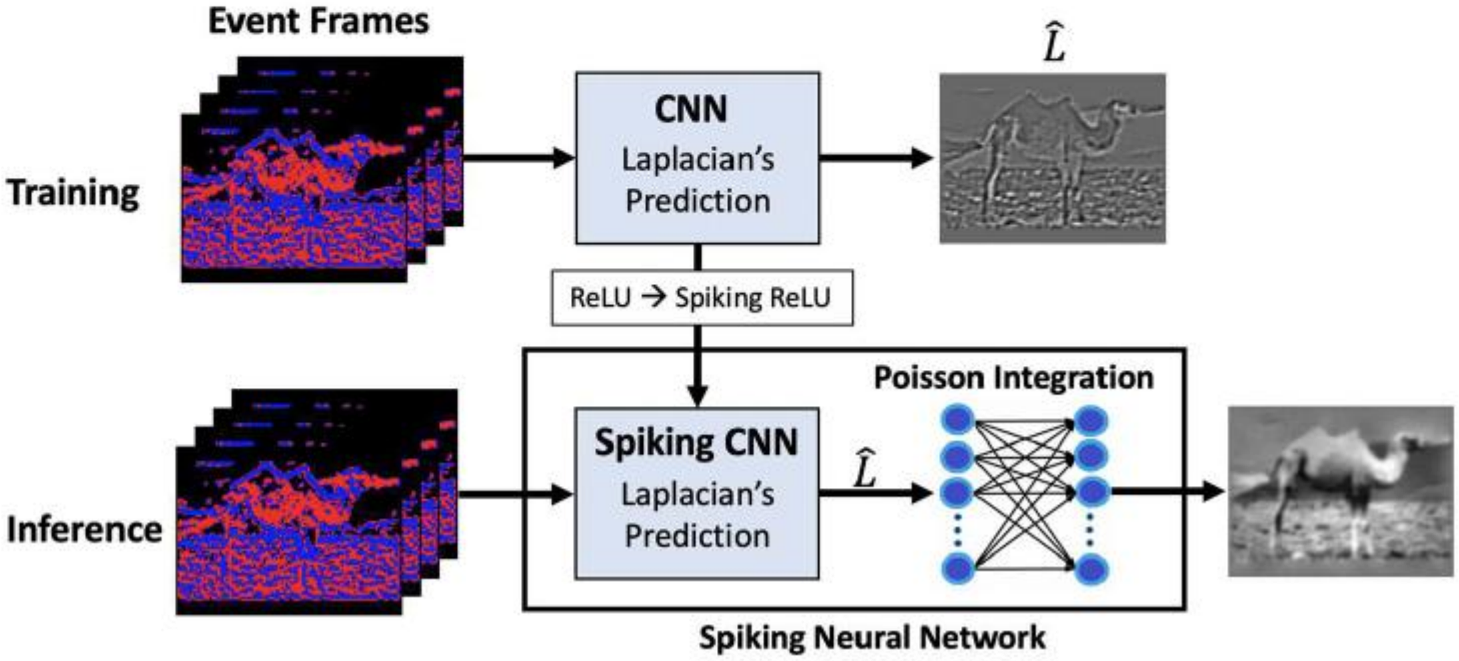
Cohen-Duwek et al. (2021) [Source 2]

Focus: Efficient image reconstruction

Model pipeline : Laplacian-Poisson

Input: Frame events → **Phase 1:** Laplacian Prediction (CNN) → **Phase 2:** Filling-in (PI) → **Output:** Image intensity

- **At training time (top):** CNNs are train to predict the Image's Laplacian. The trained CNNs then converted to SNNs for inference.
- **At inference time (bottom):** a fully spiking implementation of image reconstruction from event camera.
- **Data**: Event-camera files from N-MNIST and N-Caltech101 datasets



[Source 2]

Model configurations

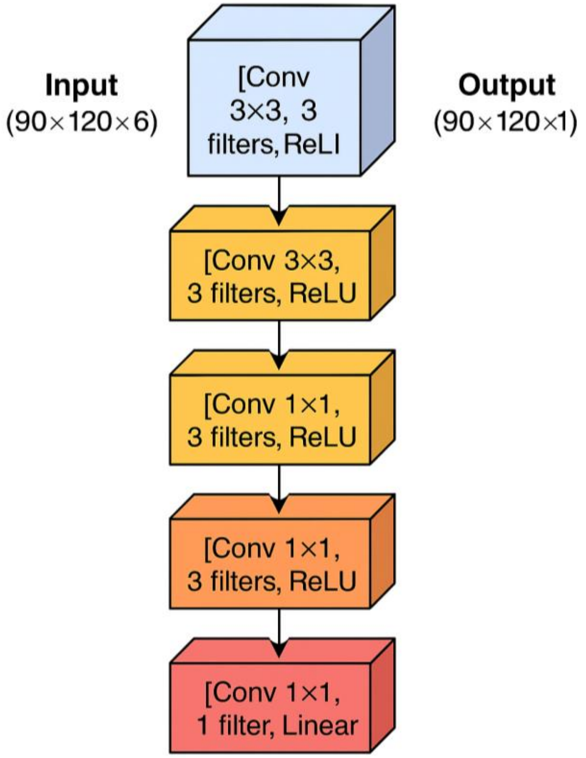
A. Two-Stage CNN→SNN Model:
   1. A compact 5-layer CNN for Laplacian prediction.
   2. SNN for Poison integration.

B. Two-Stage CNN→SNN Ultra-Lightweight Model– **NOVAL**
   1. Shared-Event Filters CNN - treats events as a video-like signal with shared filters across frames.
   2. SNN for Poison integration.

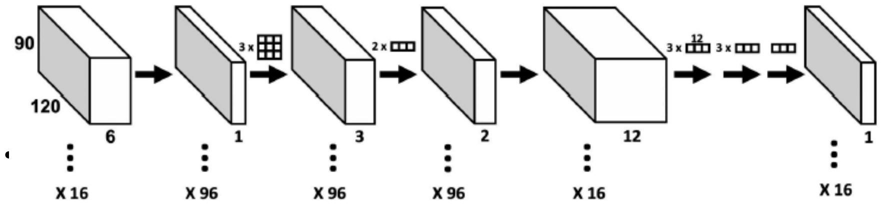
C. Two-Stage Fully Spiking CNN→SNN Model:
   1. The CNN is converted into a SNN CNN using NEF.
   2. SNN for Poison integration.

D. Direct Reconstruction CNN Model:
   1. A 5-layer CNN directly reconstructs the image without Laplacian prediction or Poisson integration.

CNN Basic architecture



Shared-Event Filters CNN (including reshape)



[Source 2]

Main Results

- **Model tested:**

A. Two-Stage CNN→SNN

Models 1-6 varied in width (number of filters).

B. Two-Stage Shared-event Filters

SM and SR : using Mish \ ReLU activation.

C. Two-Stage Fully Spiking

SNN.1 and SNN5: model#5 varied in maximal firing rate of 100 and 5,000 Hz.

D. Direct Reconstruction

Models 1,4,5 are varied in width (number of filters).

Metrics Used:

Peak Signal-to-Noise Ratio (PSNR) ↑
Structural Similarity Index Measure (SSIM) ↑
Mean Square Error (MSE) ↓

	#	filters	params	PSNR	SSIM	MSE
$\hat{I}$+PI	1	50,100,50,20,1	53,941	24.29	0.864	0.0042
	2	50,50,50,20,1	28,891	24.31	0.859	0.0042
	3	20,20,20,20,1	5,581	24.47	0.858	0.0042
	4	10,10,10,10,1	1,691	24.67	0.861	0.0039
	5	3,3,3,3,1	277	24.23	0.844	0.0042
	6	3,1,3,1,1	205	23.86	0.838	0.0047
	S_M	3*,2*,3,3,1	93	23.04	0.824	0.0058
	S_R	3*,2*,3,3,1	93	11.40	0.326	0.0734
	SNN^1	3,3,3,3,1	277	17.90	0.679	0.0217
	SNN^5	3,3,3,3,1	277	19.36	0.756	0.0147
$\hat{I}$	1	50,100,50,20,1	53,941	19.53	0.519	0.0117
	4	10,10,10,10,1	1,691	19.60	0.512	0.0114
	5	3,3,3,3,1	277	19.22	0.490	0.0124

[Source 2]

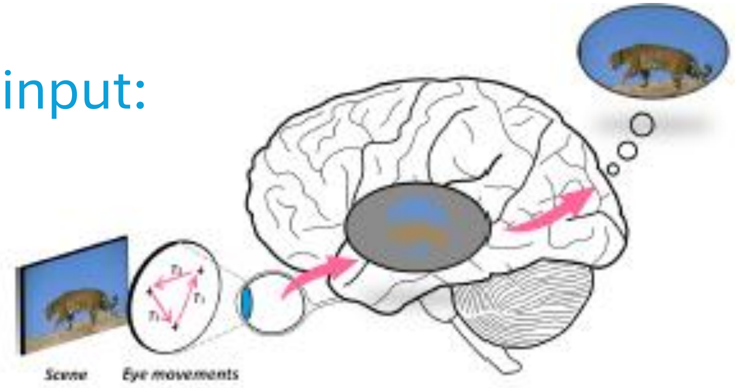
3. Expanding to Color and Peripheral Vision

Perceptual Colorization of the Peripheral Retinotopic Visual Field using Adversarial-Optimized Neural Networks

Cohen-Duwek et al. (2023) [Source 3]

Focus: Modeling Visual Perception from more realistic retinal input:

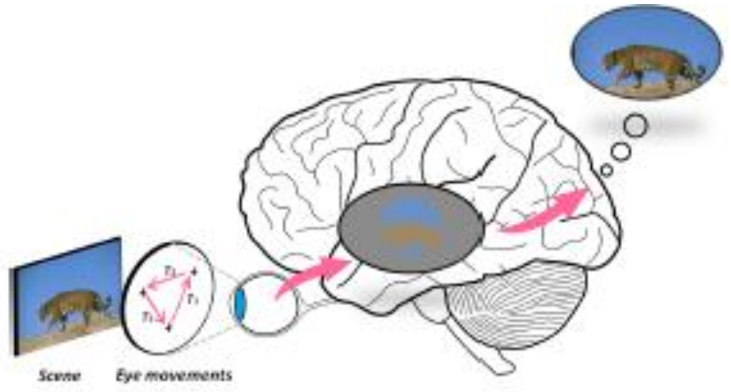
Peripheral vision, achromatic cues, and saccades



Methods

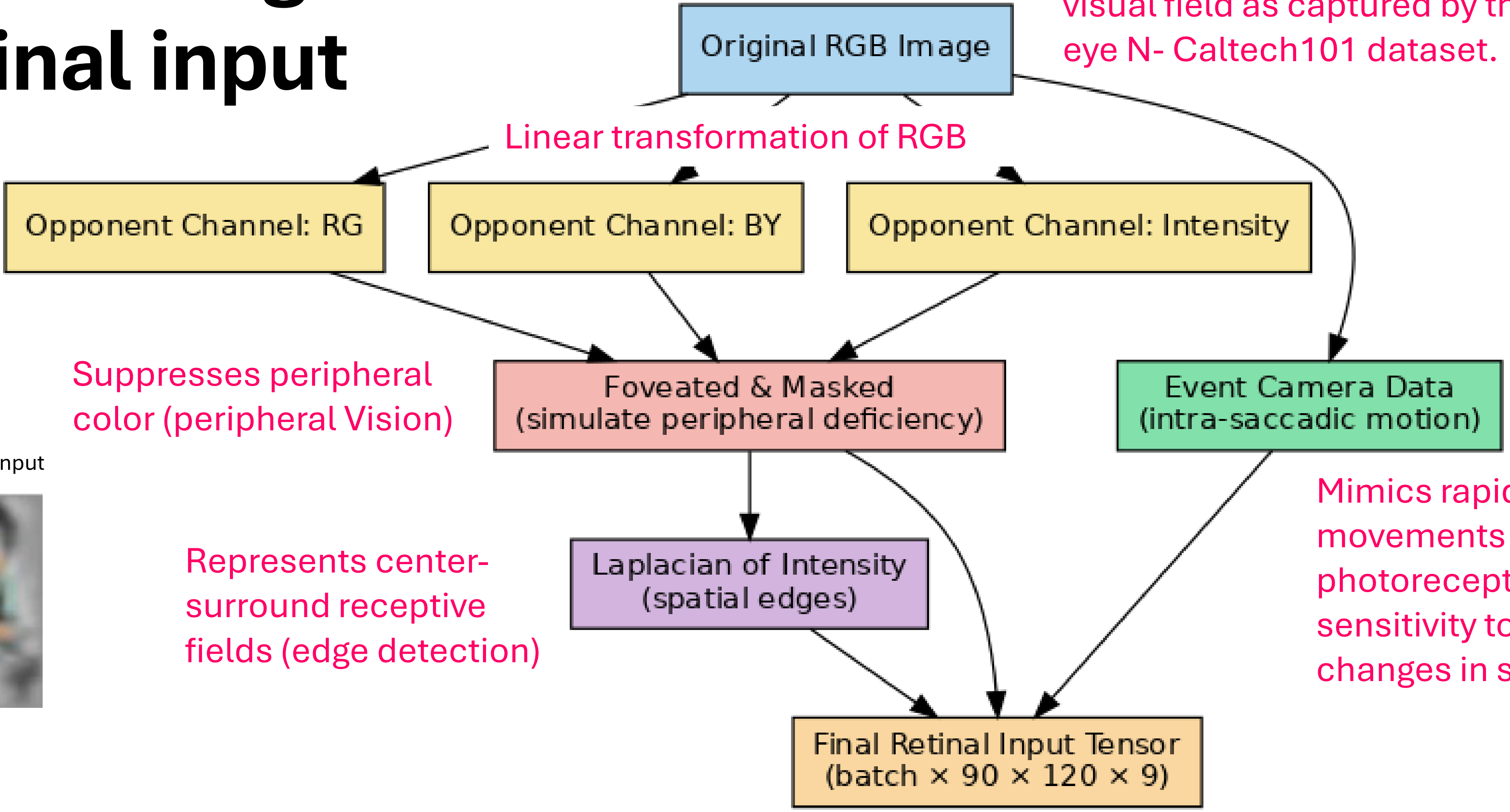
- Computational Goal:
  - Reconstruct high-resolution color images from limited retinal input:

1. Generate synthetic retinal inputs:
   - Opponent colors and intensity channels
   - reduced peripheral color
   - Event camera frames (intra-saccadic motion)
2. Reconstruction:
   - U-Net architecture for image colorization
   - Multi–stage GAN trained end-to-end



[Source 3]

Generating retinal input



visual field as captured by the eye N- Caltech101 dataset.

Linear transformation of RGB

Mimics processing in ganglion and retinal cells.

Suppresses peripheral color (peripheral Vision)

Represents center-surround receptive fields (edge detection)

Mimics rapid eye movements and photoreceptors sensitivity to changes in stimuli

9 channels = RG, BY, Intensity + 6 event frames

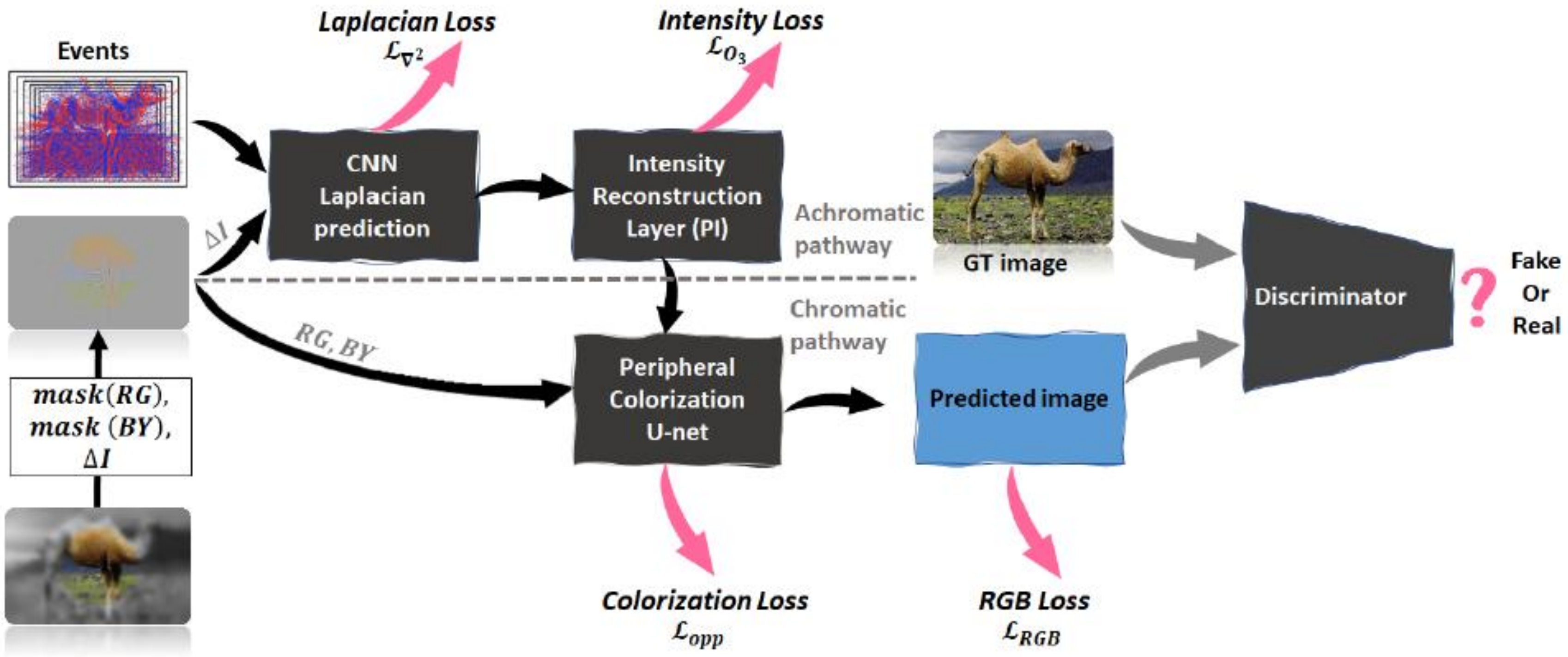
Foveted retinal input

[Source 3]

The architecture of the reconstruction and colorization model

- Proposed a sophisticated Generator:
  1. CNN for Laplacian prediction (source 2).
  2. Poisson Solver for intensity reconstruction
  3. U-Net for image colorization
- Trained end-to-end using an adversarially trained Discriminator (PatchGAN)

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \mathcal{L}_{RGB} + \mathcal{L}_{opp} + \mathcal{L}_{O_3} + \mathcal{L}_{\nabla^2}$$



[Source 3]

Models Config. & Main Results

Method	SSIM ↑	LPIPS ↓
Events + D	0.7840	0.2120
Events - D	**0.8329**	**0.1643**
No Events + D	0.7651	0.2240
No Events - D	0.7682	0.2114

- **Model tested**:
  - **Classical:** Minimize Generator losses (MAE for Laplacian & colors, SSIM/LPIPS for similarity).
  - **Adversarial (GAN):** Minimize combined Generator + GAN loss.
- **Results:**
- Both methods reconstruct high-quality images from incomplete retinal input.
- **Adversarial training:** Produces more colorful peripheries (often green)) , though sometimes less consistent.
- **Event data:** Produces sharper peripheries and improves SSIM & LPIPS.
- **Best model:** Event data + adversarial training
- → most vivid peripheral colorization, closest to human perception (predicting/filling in missing colors).

[Source 3]

4. Achieving Visual Stability in Active Vision

Reconstruction of Visually Stable Perception from Saccadic Retinal Inputs Using Corollary Discharge Signals-Driven ConvLSTM Neural Networks

Cohen-Duwek et al. (2024) [Source 4]

Focus: Visually Stable Perception

Methods

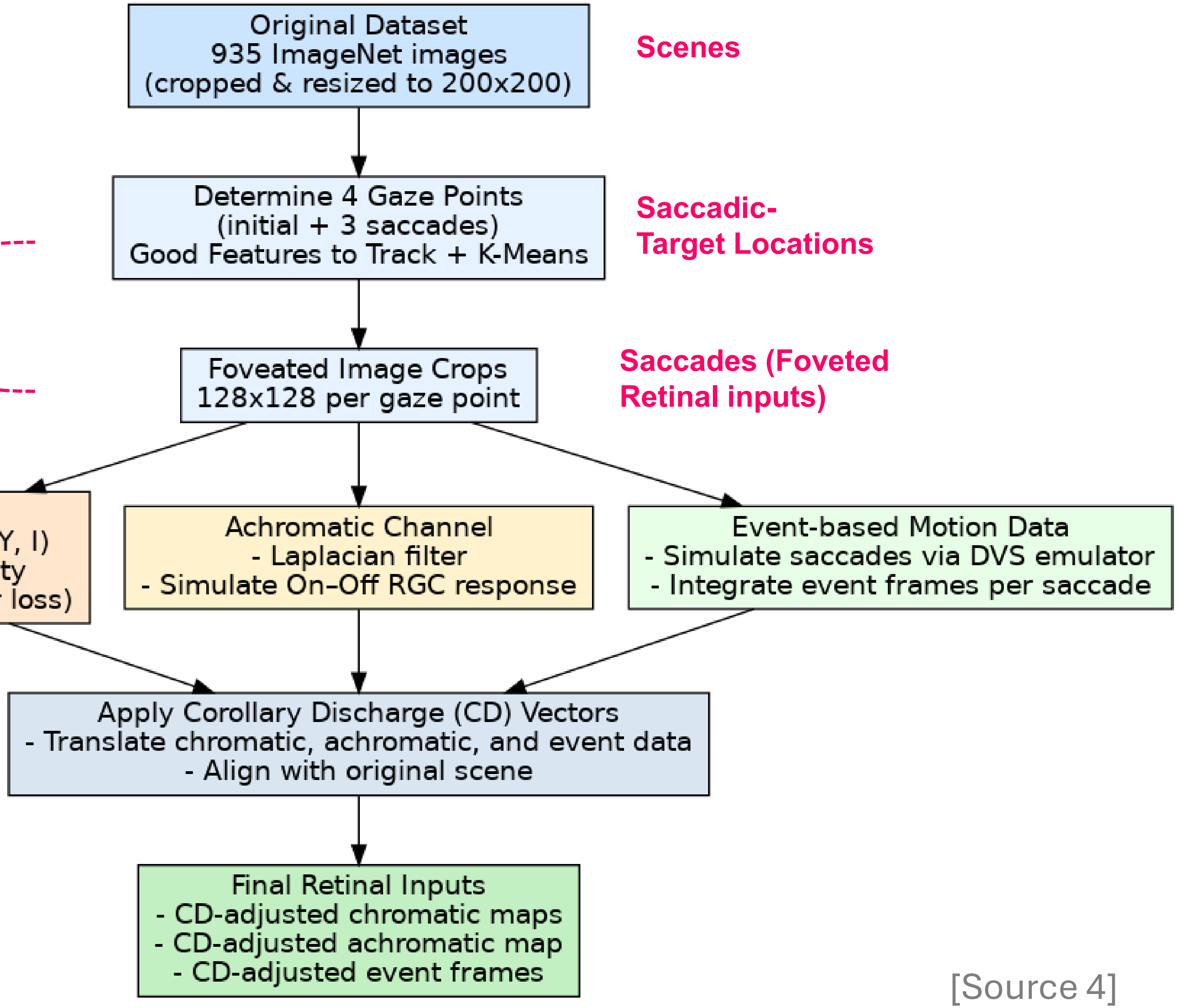
- Computational Goal:
  - Reconstruct a visually stable, high-resolution color image from dynamic saccadic inputs

1. Enhance dataset for Active Vision:
   - Dynamic saccades motion (towards salient features within image)
   - CD signals (anticipatory adjustments by translation vectors)
2. Multi-phase ConvLSTM network :
   - Functional visual memory by trans-saccadic Integration (predictions from each successive saccade)

[Source 4]

Generating retinal input

Dynamic Saccades:
Simulated eye movements towards **points of interest**



Scenes

Saccadic-Target Locations

Saccades (Foveted Retinal inputs)

For each saccade:

Foveated Retinal Inputs

[Source 4]

The architecture of the reconstruction and colorization of stable images

- GAN-based Architecture with End-to-end minimization (combined loss function across components)
- The Generator functions as a multi-stage network:

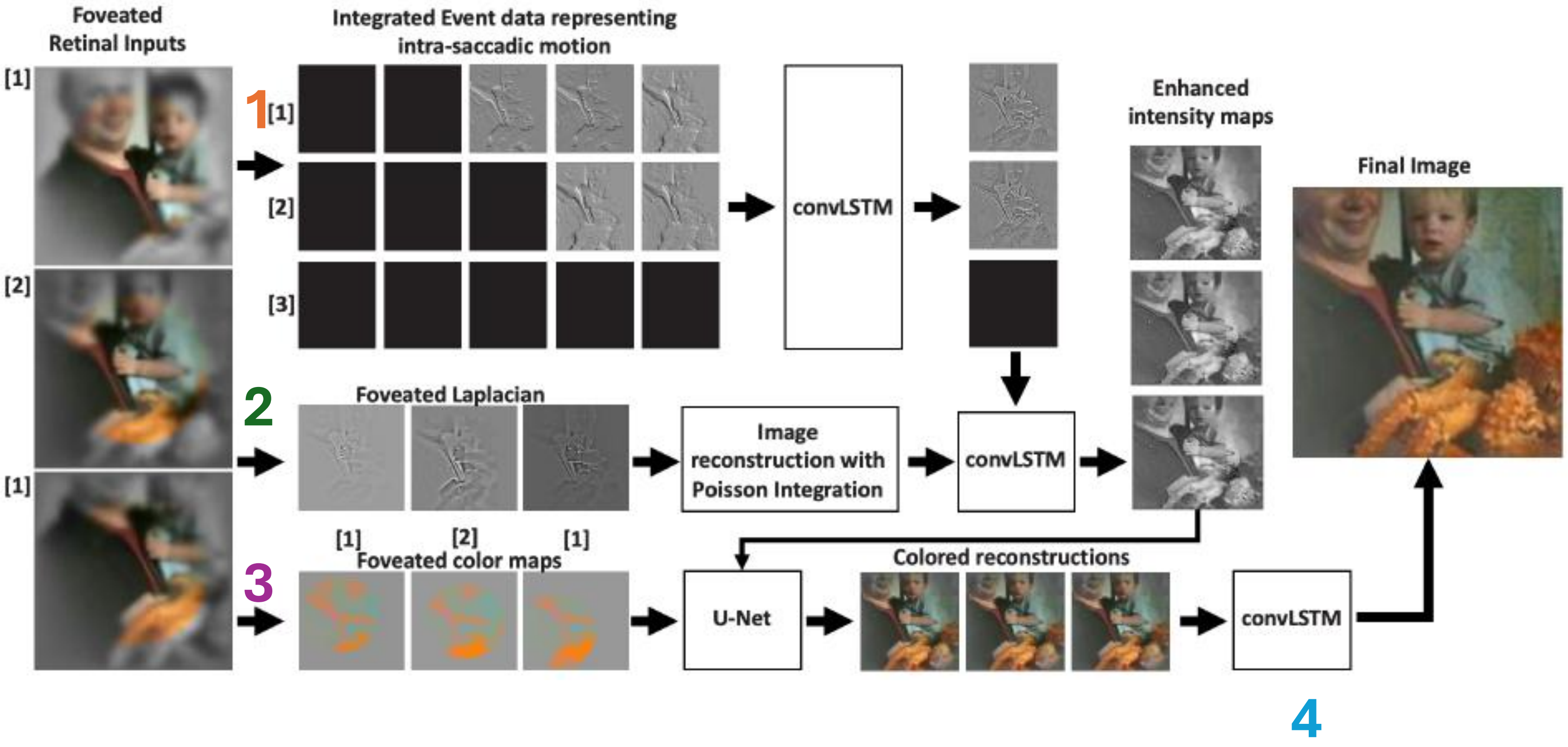
for each saccade:

1. Initial Intensity Reconstruction (ConvLSTM from event frames)

2. Intensity Prediction and Enhancement (ConvLSTM)

3. Colorization (U-Net)

4. Saccadic Integration for final Stabilization (ConvLSTM)

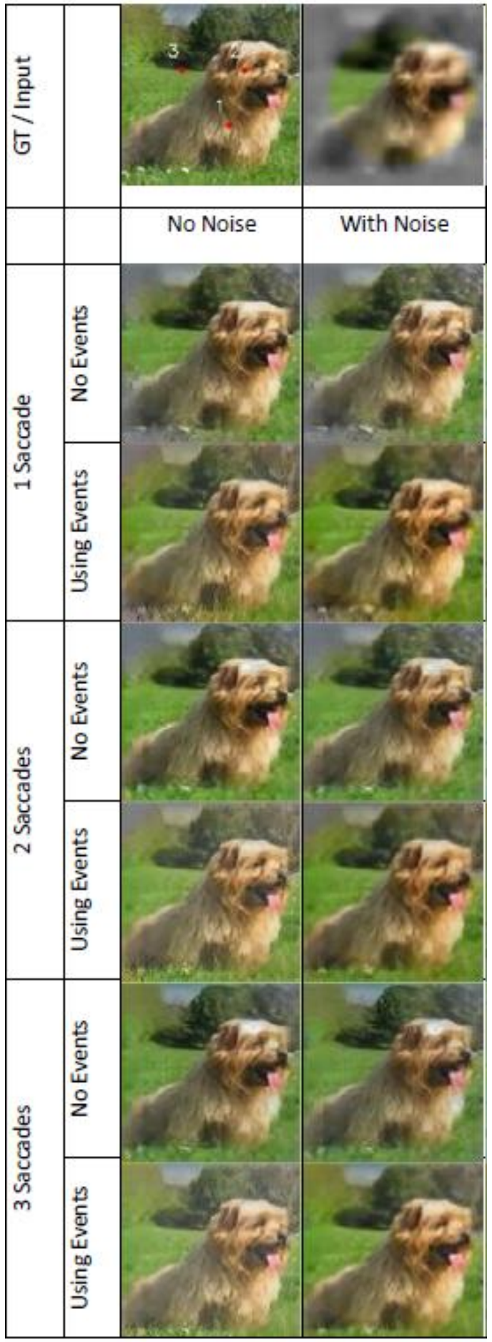


[Source 4]

Main Results

		1 Saccade				2 Saccades				3 Saccades			
		No Events		With Events		No Events		With Events		No Events		With Events	
	Input	No Noise	With Noise	No Noise	With Noise	No Noise	With Noise	No Noise	With Noise	No Noise	With Noise	No Noise	With Noise
SSIM (%)	59.42	75.62	71.65	77.86	73.75	78.39	73.17	81.05	75.08	80.63	73.68	**82.78**	75.88
LPIPS (%)	53.54	30.32	31.69	27.35	31.21	26.22	28.01	24.13	28.52	22.92	25.27	**21.73**	26.67
PSNR (dB)	17.01	23.22	22.74	23.57	23.24	23.72	23.08	24.04	23.5	**25.19**	24.15 dB	25.17	24.39 dB
CIEDE2000	13.72	6.51	6.92	6.52	6.7	6.26	6.59	6.10	6.38	**5.3**	5.75	5.31	5.66

- More saccades → better quality: reconstructions became sharper and more colorful, (higher SSIM & LPIPS) despite lower pixel accuracy.
- Event-based inputs improved perceptual similarity (higher SSIM & LPIPS), though PSNR and CIEDE2000 were sometimes better without events.
- CD signals are essential for stable perception - Adding Gaussian noise to CD vectors degraded all metrics and caused visual blurriness.



[Source 4]

Overall Conclusion and Future Directions

- **Future Directions & Ongoing Work**
  - Demonstrate improvement in image reconstruction with more than three consecutive saccades
  - Increase realism by modeling an imperfect CD signals
- **Bridging the Perception Gap — Conclusions**
  - Robust progression in computational models for visual perception and realism
  - Brain → model : Principles observed in the brain can be computationally realized
  - Model reproduces Cohen et al. (2020) experiments
  - Neuromorphic computation for perceptual filling-in (SNNs,NEF, event-frame data)

Perceptual Filling-In Module

Efficient Image Reconstruction

Perceptual Realism

Perceptual Stability in Active Vision

Implementations:

Introduction and Target Model

Methodology and Experimental Setup

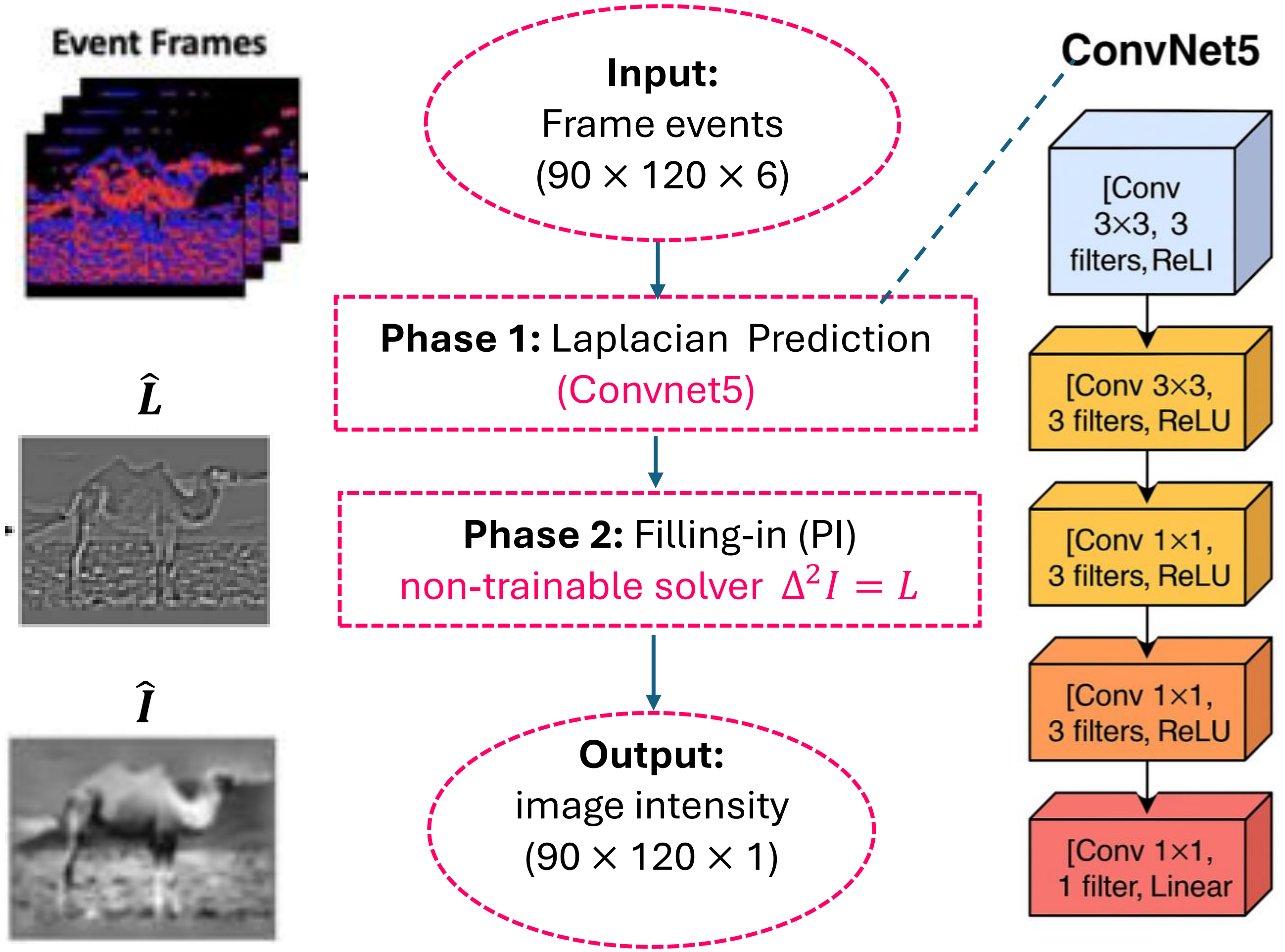
Results and Conclusions

Introduction and Target Model

- **Goal:** Sensitivity Study
- **Target Model:** ConvNet5 (Model #5): Lightweight design of event-Based Image Reconstruction [Source 2] .
- **Params:** ~277
- **Composite Loss =**

$$\lambda_1 MAE(L, \hat{L}) + \lambda_2 \left(1 - SSIM\left(PI(L), PI(\hat{L})\right)\right) + \lambda_3 Edge\_Loss\left(PI(L), PI(\hat{L})\right)$$

- **Training setup**: Adam+ early stopping on validation SSIM.



Methodology and Experimental Setup

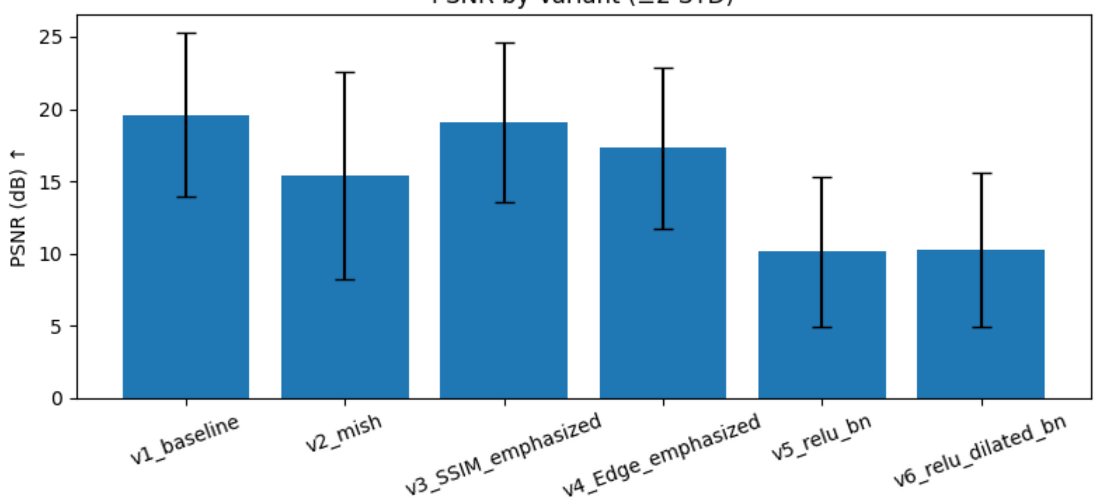
- **Dataset**: N-Caltech101
- **Training Method**: Same setup as original
- **Initialization:** Use pretrained weights (scratch training only for architecture variants)
- **Training Data**: Limited subset -150 samples, batch size – 16, epoch number – varied 10-50

Variant Category	Specific Changes Tested	Purpose
Activation Function	Replacing ReLU with Mish activation.	Test if Mish (a smooth, non-monotonic function) improves quality.
Loss Weighting	SSIM Emphasis or Edge Emphasis (adjusting λ)	Examine effects on perceptual smoothness or contour sharpness.
Architectural Changes	Adding Batch Normalization (BN), with or without Dilation.	Expected to improve stability or expand the receptive field.

Results and Conclusions

Performance Comparison of Variants

Experiment	Params	PSNR ↑	SSIM ↑	MSE ↓	PSNR CV	SSIM CV	MSE CV
Baseline	277	19.599	0.761	0.013885	0.144	0.129	0.881818
Mish	277	15.380	0.670	0.041912	0.233	0.184	1.081851
SSIM Emph.	277	19.073	0.748	0.015453	0.146	0.134	0.839801
Edge Emph.	277	17.301	0.710	0.022825	0.161	0.147	0.703932
ReLU+BN	325	10.135	0.360	0.116998	0.257	0.391	0.683272
ReLU+Dil.+BN	325	10.235	0.381	0.115278	0.262	0.377	0.687837



$$\lambda_{Baseline} = (1.0, 0.25, 0.25)$$
$$\lambda_{SSIM\ Emph.} = (1.0, 0.35, 0.15)$$
$$\lambda_{Edge\ Emph} = (1.0, 0.15, 0.35)$$

Results and Conclusions

Visual Performance Comparison of Variants

Discussion

- Lightweight ConvNets are **highly sensitive** to architectural perturbations.
- Initialization and training stability are **critical** under limited data.
- Mish or SSIM emphasis followed qualitative expectations but not quantitatively stronger.
- BatchNorm/dilation destabilized learning maybe due to model size constraints.

References

- Biologically Plausible Spiking Neural Networks for Perceptual Filling-In Cohen Duwek and Tsur (2021) [Source 1]
- Image Reconstruction from Neuromorphic Event Cameras using Laplacian-Prediction and Poisson Integration, Cohen-Duwek et al. (2021) [Source 2]
- Perceptual Colorization of the Peripheral Retinotopic Visual Field using Adversarially-Optimized Neural Networks, Cohen-Duwek et al. (2023) [Source 3]
- Reconstruction of Visually Stable Perception from Saccadic Retinal Inputs Using Corollary Discharge Signals-Driven ConvLSTM Neural Networks, Cohen-Duwek et al. (2024). [Source 4]
- The limits of color awareness during active, real-world vision (Cohen et al. 2020)