

# A Deep Learning Approach for Image-based Semantic Segmentation with Preserved Interpretability

Juan Carlos Aguirre Arango  
[jucaguirrear@unal.edu.co](mailto:jucaguirrear@unal.edu.co)



Universidad Nacional de Colombia  
Signal Processing and Recognition Group - SPRG  
Advisor: Andrés Marino Álvarez Meza, Ph.D.  
Co-advisor: César Germán Castellanos Domínguez, Ph.D.

September 20, 2023



# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

7 Results

8 Conclusions

9 References



# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

7 Results

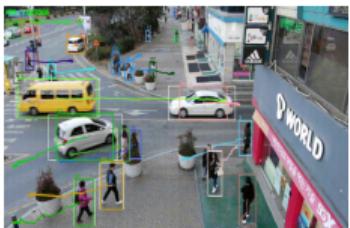
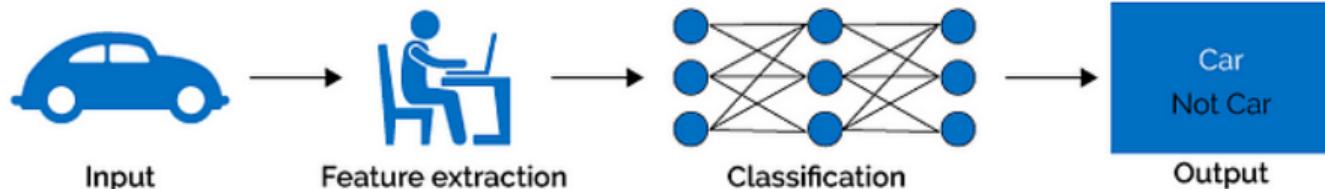
8 Conclusions

9 References

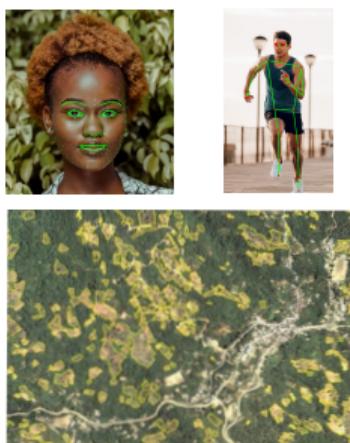
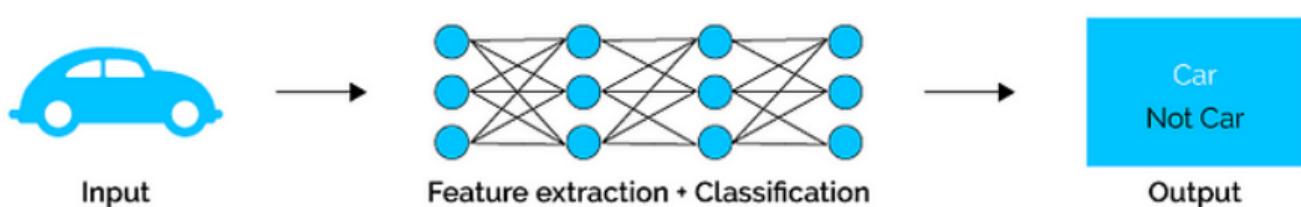


# Classic Machine Learning vs. Deep Learning

## Classic Machine Learning

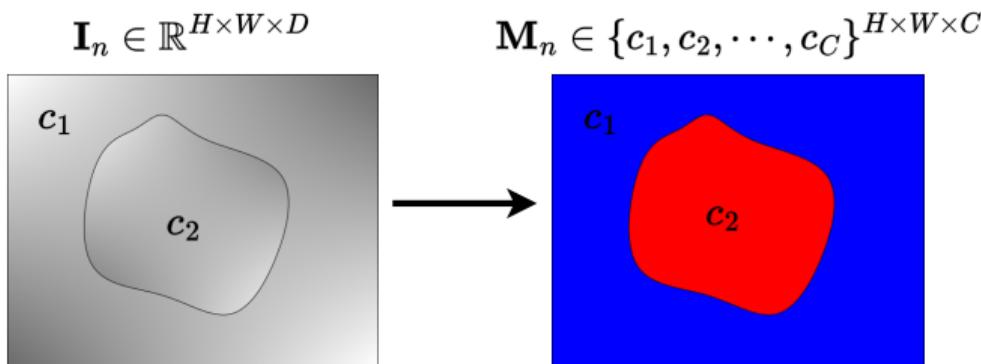


## Deep Learning





# Semantic Segmentation (SS)



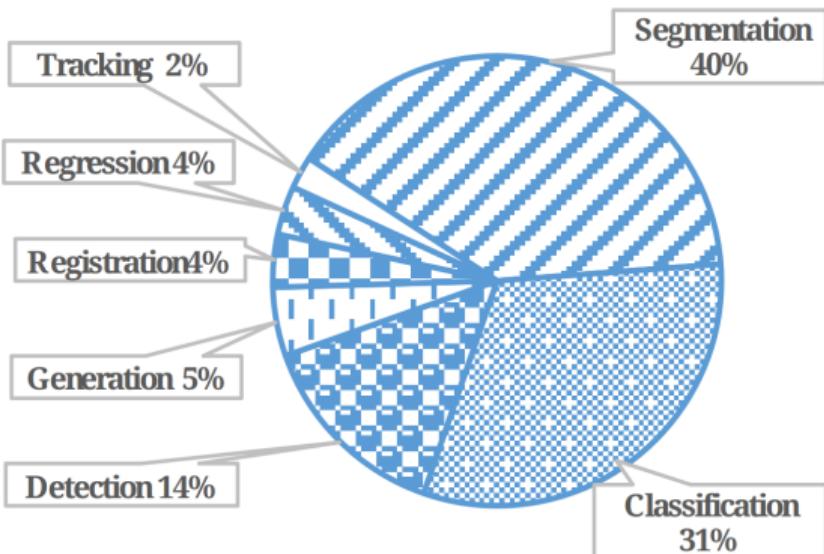
Deep learning models have allowed the use of semantic segmentation in multiple tasks [Alokasi and Ahmad, 2022]

- Autonomous Vehicles [Cakir et al., 2022, Tsai et al., 2023, Rizzoli et al., 2022, Burel et al., 2022]
- Virtual Reality [Tiator et al., 2020, Lin et al., 2016]

- Health care [Qureshi et al., 2023, Khan et al., 2021, Soomro et al., 2023]
- Agriculture [Luo et al., 2023, Anand et al., 2021]
- Satellite Imagery [Lilay and Taye, 2023, Wieland et al., 2023]



# Main Tasks in Medical Image Analysis [Li et al., 2021]



Most diagnosis methods rely on changes in morphological information of the region of interest [Kriti et al., 2022].

Small sample sizes and heterogeneous regions of interest are presented in medical image

[Castiglioni et al., 2021].



# Signal Processing and Recognition Group - SPRG

- The SPRG is interested in developing machine learning systems to analyze biosignals [Bron et al., 2015, Jimenez et al., 2018, Jimenez-Castaño et al., 2021].
- Nowadays, the SPRG is developing objective and low-cost monitoring procedures for obstetric patients under regional neuraxial analgesia.



**S.E.S.**  
Hospital de Caldas

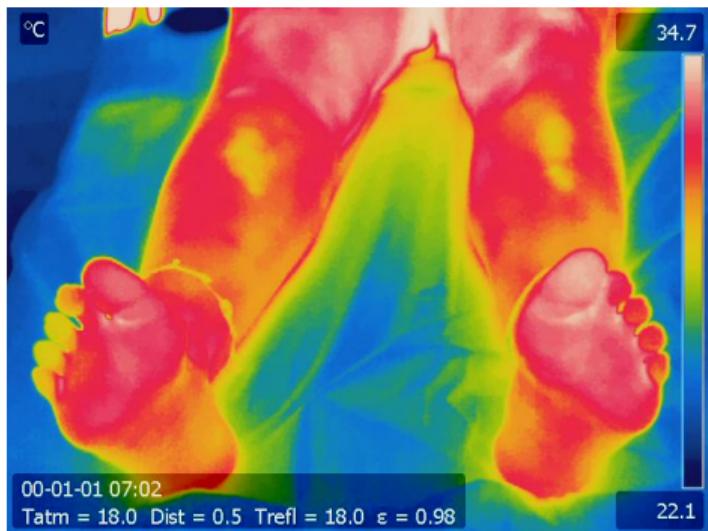
**UTP**  
Universidad Tecnológica  
de Pereira





# Childbirth Pain Monitoring I

Thermal imaging is an objective and non-invasive technique to quantify warm modifications after catheter placement due to blood flow redistribution [Haren et al., 2013].

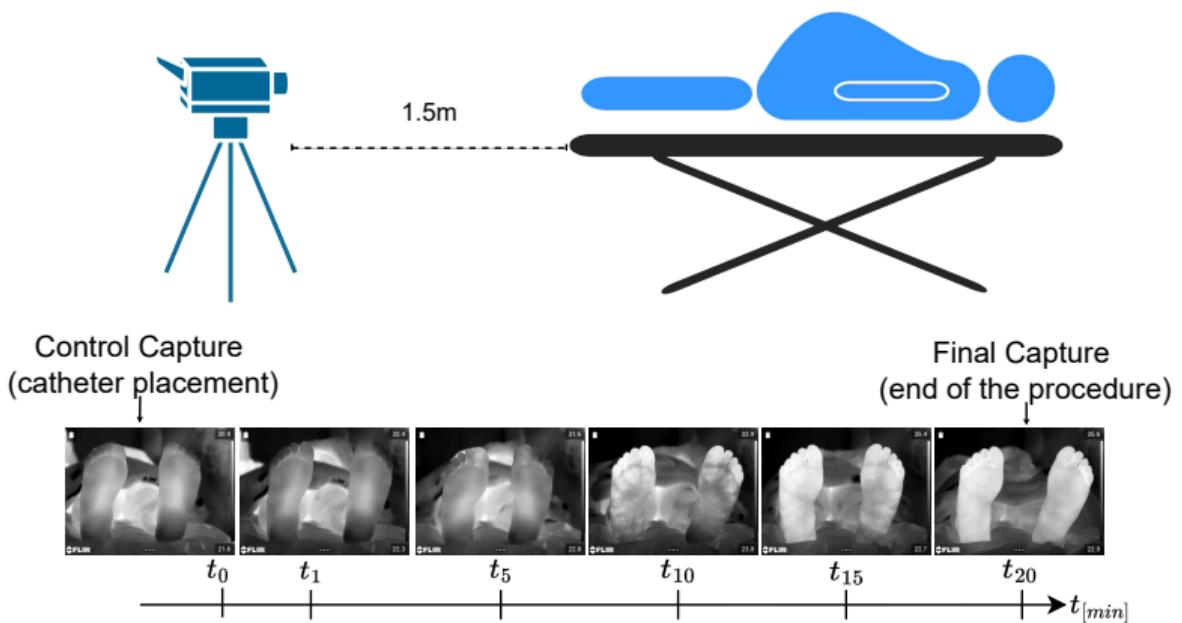


**Reliable and automatic thermal measurements are required during childbirth analgesia monitoring** [Asghar et al., 2014].

Figure: Thermal imaging example.



# Childbirth Pain Monitoring II



Medical specialists from SES Hospital de Caldas designed a protocol of acquisition



# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

7 Results

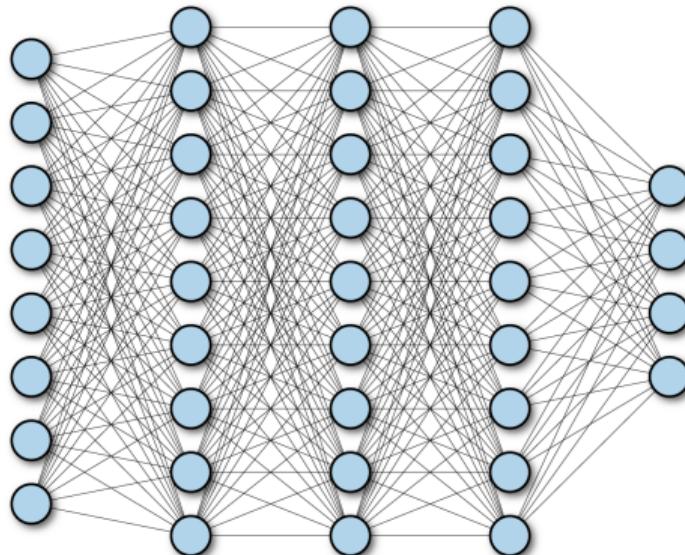
8 Conclusions

9 References



# Issue #1: Small Sample Size and Overfitting

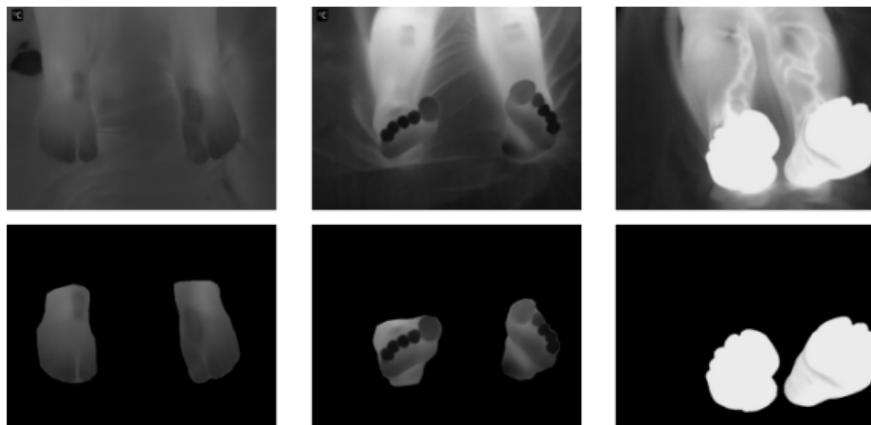
- Challenges in acquiring sufficient datasets in obstetrics environments [Melesse et al., 2022, Willemink et al., 2020]
- Effective training requires extensive annotations [Sarker, 2021, Li et al., 2021].





## Issue #2: High Variability in the ROI

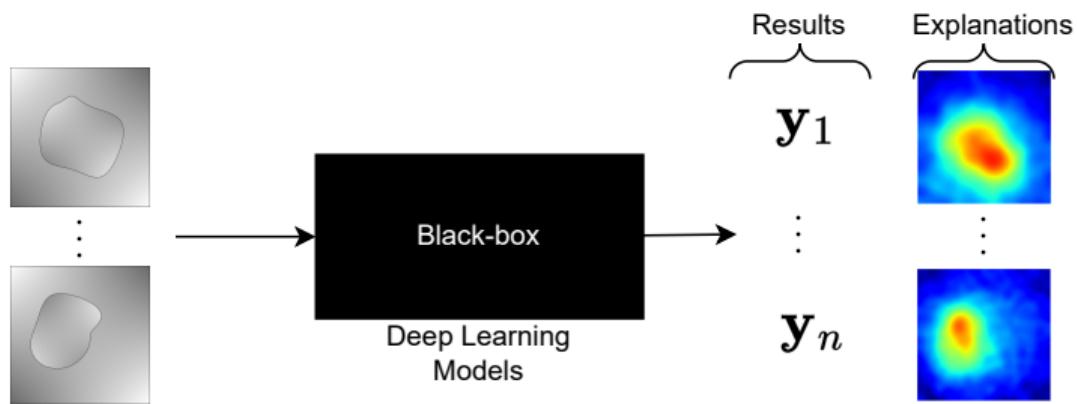
- Variations in anatomy, pathology, and imaging parameters result in significant variations in the ROI's shape, size, and texture [Li et al., 2021].
- Diverse ROI positions can lead to images with different orientations, sizes, and shapes, even within the same subject. This variability includes cases where the ROI may overlap or be partially obscured [Arteaga-Marrero et al., 2021].





## Issue #3: Lack of Quantitative Measures for Interpretability in SS Models

- Deep learning-based semantic segmentation models are often black-boxes, making their explanations challenging [Linardatos et al., 2020].
- Current methods for explainability rely mainly on visual inspection or qualitative analysis, limiting the evaluation of model performance [Wang et al., 2022, Salahuddin et al., 2022].





# Research Question

How can we develop a method for generating **local and equivariant representations** that can improve the **generalization of deep-learning** models while maintaining **interpretability** in **Semantic Segmentation** tasks for medical images?



# Outline

- 1 Motivation
- 2 Problem Statement
- 3 Literature Review
- 4 Aims
- 5 Methodology
- 6 Experimental Set-up
- 7 Results
- 8 Conclusions
- 9 References



# Small Sample size and Overfitting

Small Sample  
Size and  
Overfitting

## Regularization techniques

- L1 and L2 regularization [Goodfellow et al., 2016]
- Batch normalization [Goodfellow et al., 2016]
- Dropout [Chen et al., 2020b]
- Early stopping [Goodfellow et al., 2016]
- Transfer learning [Alzubaidi et al., 2020]
- Data Augmentation [Shorten and Khoshgoftaar, 2019]

Reduction of complexity  
decreases the capability of  
the models to learn common  
dependencies

## Architecture enhancement

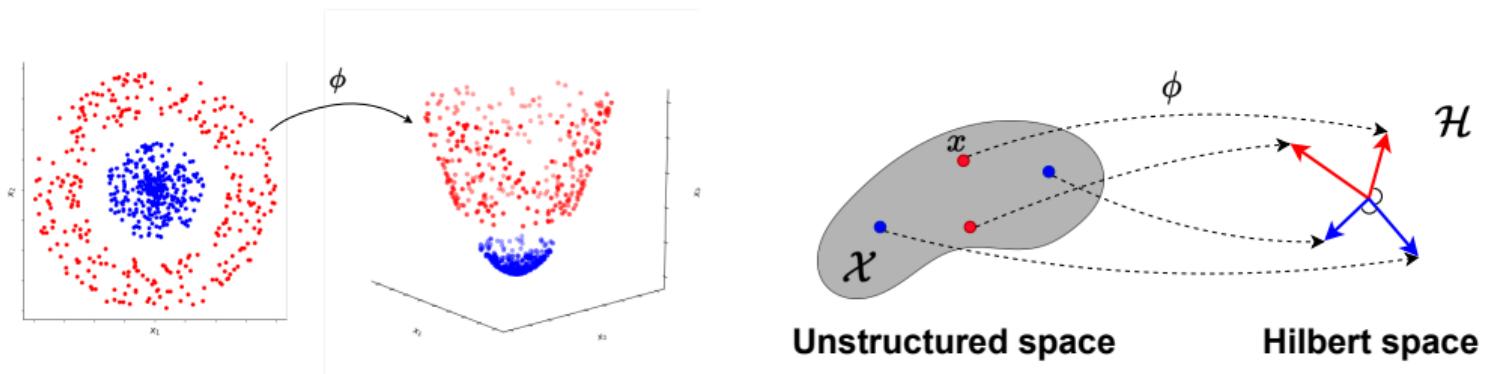
- Cosine-CKN [Mohammadnia-Qaraei et al., 2018]
- ConvRFF with Bayes [Wang et al., 2021]
- ConvRFF Bayes and bypass [Wang et al., 2021]
- RFF U-Net-like [Jimenez-Castaño et al., 2021]

Lack of semantic segmentation  
models

Random Fourier features bring properties of kernel methods within deep learning frameworks



# Kernel Methods



- Kernel trick:  $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}}$  where  $\phi : \mathcal{X} \rightarrow \mathcal{H}$
- Mercer condition:  $\sum_{i=1}^n \sum_{j=1}^n k(\mathbf{x}_i, \mathbf{x}'_j) c_i c_j \geq 0$
- Reproducibility:  $f(\mathbf{x}) = \sum_{n=1}^N \alpha_n k(\mathbf{x}, \mathbf{x}_n) = \langle \boldsymbol{\omega}, \phi(\mathbf{x}) \rangle_{\mathcal{H}}$

It is difficult to scale kernel methods to large datasets



# Bochner's Theorem and Random Fourier Features (RFF)

A continuous function of the form  $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$  is positive definite if and only if  $k(\delta)$  is the Fourier transform of a non-negative measure.

$$k(\mathbf{x} - \mathbf{x}') = \int_{\mathbb{R}^P} p(\omega) \exp(j\omega^\top (\mathbf{x} - \mathbf{x}')) d\omega = \mathbb{E}_\omega \left\{ \exp(i\omega^\top \mathbf{x}) \exp(-i\omega^\top \mathbf{x}') \right\}$$

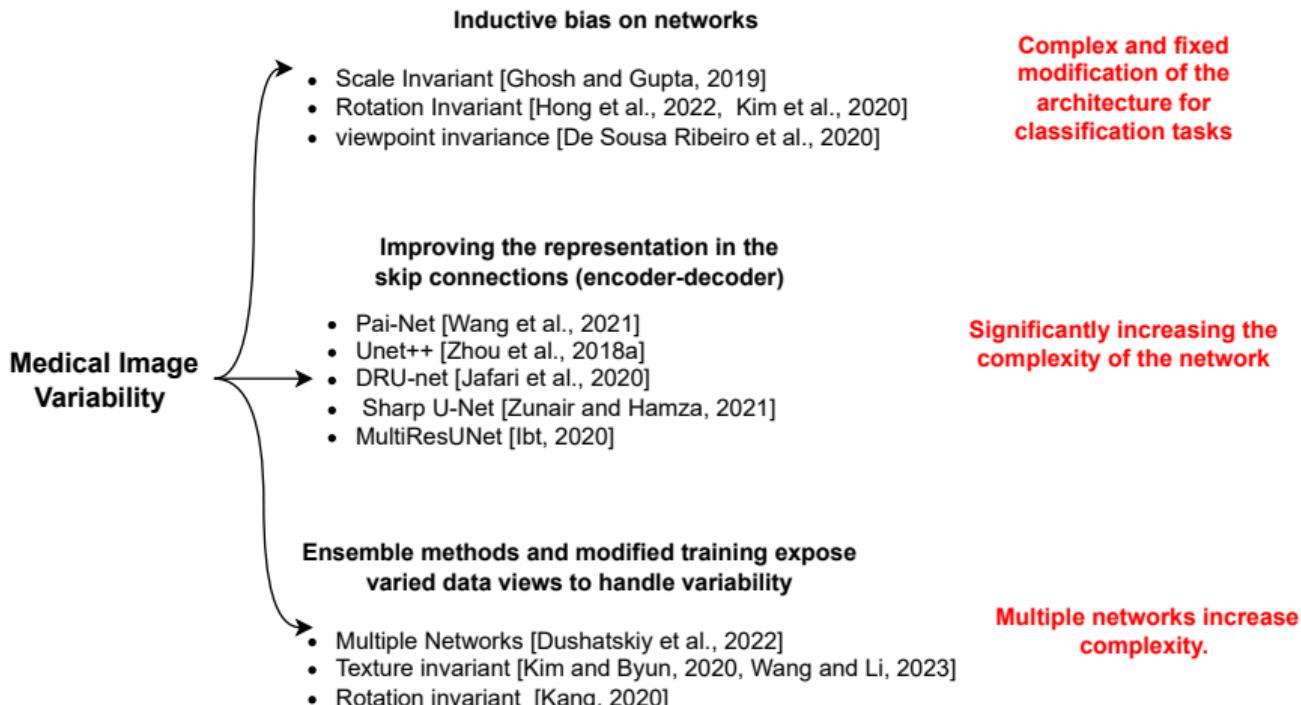
Where  $\omega \sim p(\omega)$

Allow us to find a representation where we can sample and get an approximation, named the RFF, for the mapping function in low dimensional space

However, the RFF is limited for 1-D data, and if it is extended to 2-D, its parameters are not optimized through gradient descent



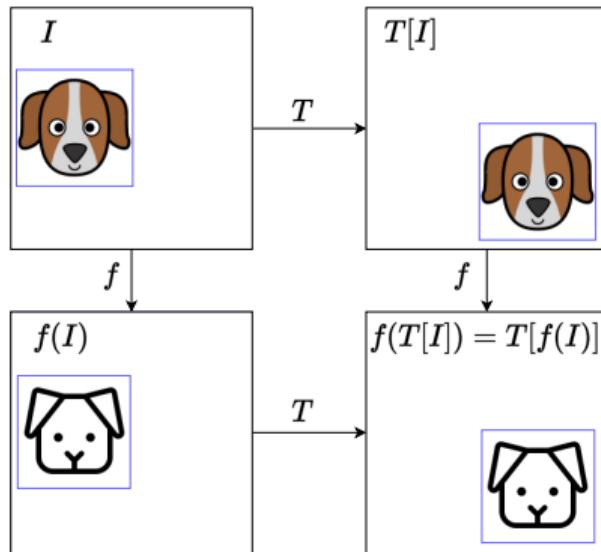
# Characterization of highly variable object patterns in CNN



Encoder-decoder architectures allow capturing both context and location information



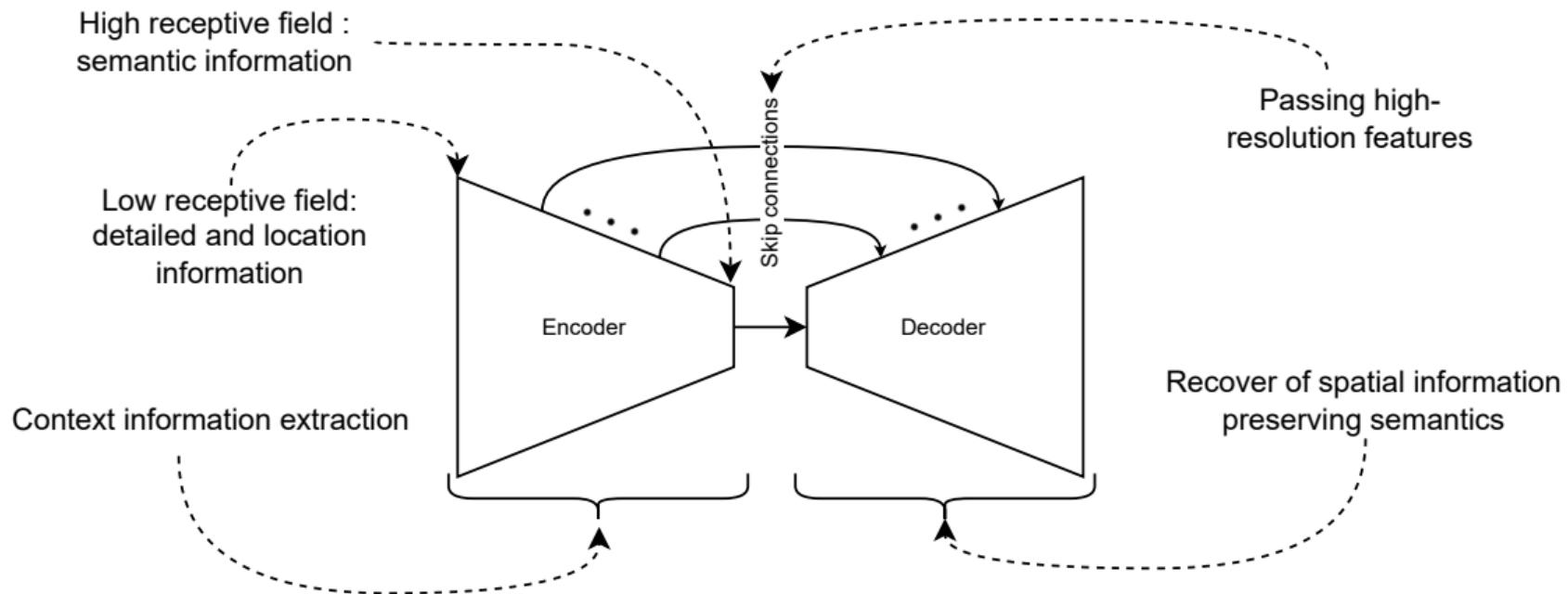
# Encoder-Decoder: Image-based Segmentation I



Translation equivariance and local properties of Convolutional layers make them efficient



# Encoder-Decoder: Image-based Segmentation II

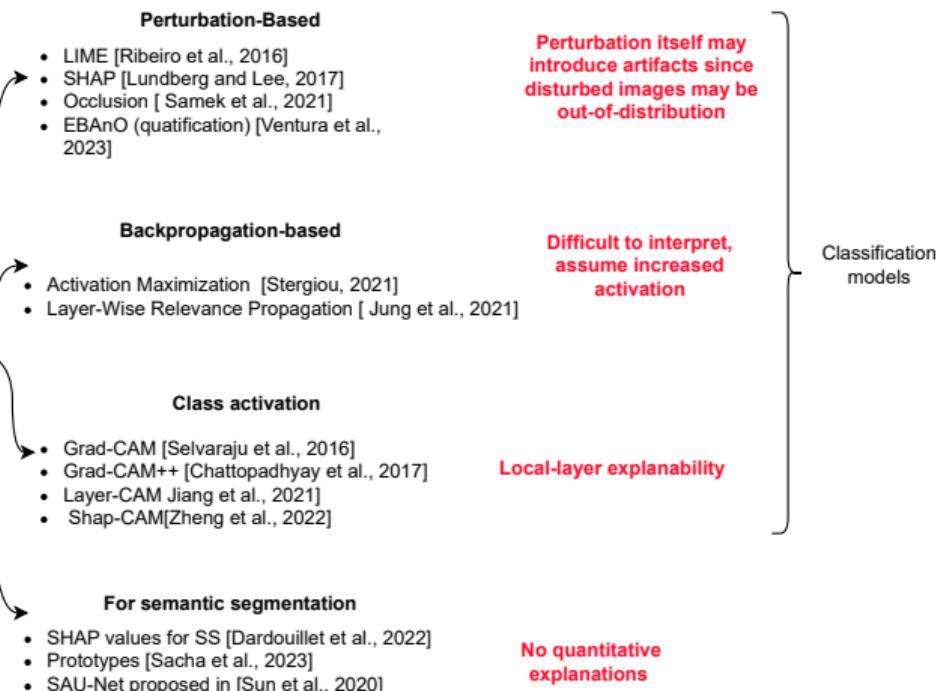


Mismatch characteristics between encoder and decoder [Zhou et al., 2020, Wang et al., 2021]



# Quantitative Interpretability of SS models

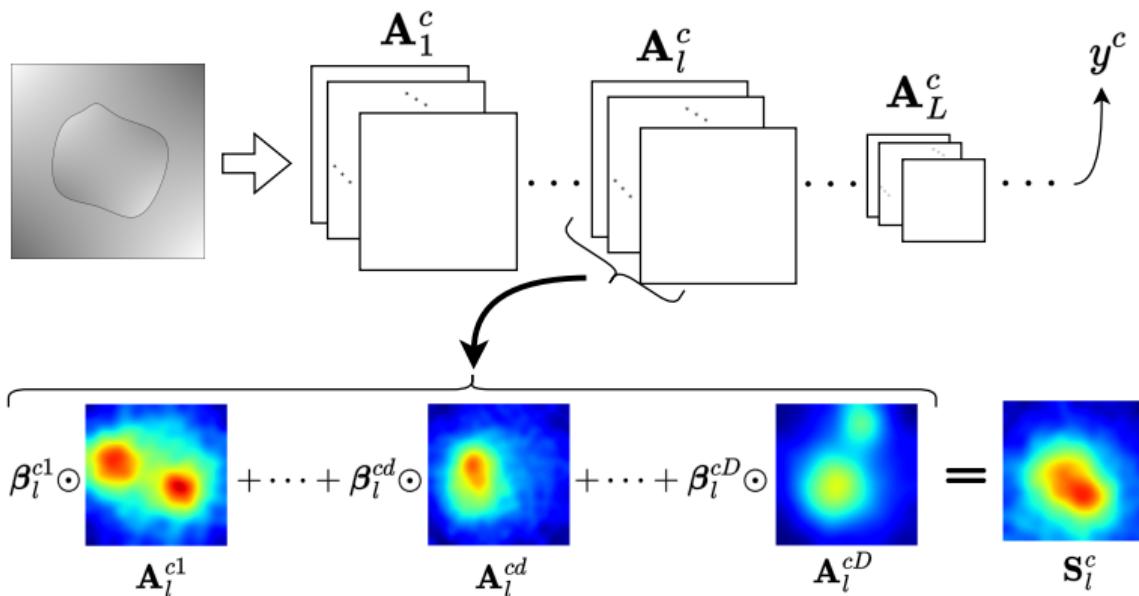
Lack of systematic and quantitative evaluations of Interpretability in Semantic Segmentation Models



Still, quantitative measures for assessment interpretability are necessary



# Class Activation Maps



CAMs provide visual inspection result analysis, with a primary focus on classification models that do not incorporate location-mask information

$\mathbf{A}$  and  $\beta$  are the feature maps, and the CAMS weights, respectively,  $\odot$  stands for Hadamard product.



# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

7 Results

8 Conclusions

9 References



# General Aim

Develop a deep-learning-based semantic image-segmentation methodology incorporating a **convolutional** layer based on **Random Fourier Features** and comprehensive **interpretability measures** to encode high variable and relevant patterns related to the region of interest and improve generalization performance under conditions of **scarce data**.



# Specific Aims

- To design an extension of **Random Fourier Features** for **spatial data** with optimization through **gradient descent** for generalization under **scarcity data** through **local and equivariant characterization**.
- To develop a **semantic segmentation** approach based on **encoder-decoder architectures** that incorporate Random Fourier Features for **enhanced skip representation** for improved capture of **small and variable** objects in semantic segmentation tasks.
- To develop a post-hoc **interpretability** approach based on measures for **quantitative assessment** of relevance maps taking into account the spatial information of semantic segmentation tasks for **global and layer-wise** relevance analysis.



# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

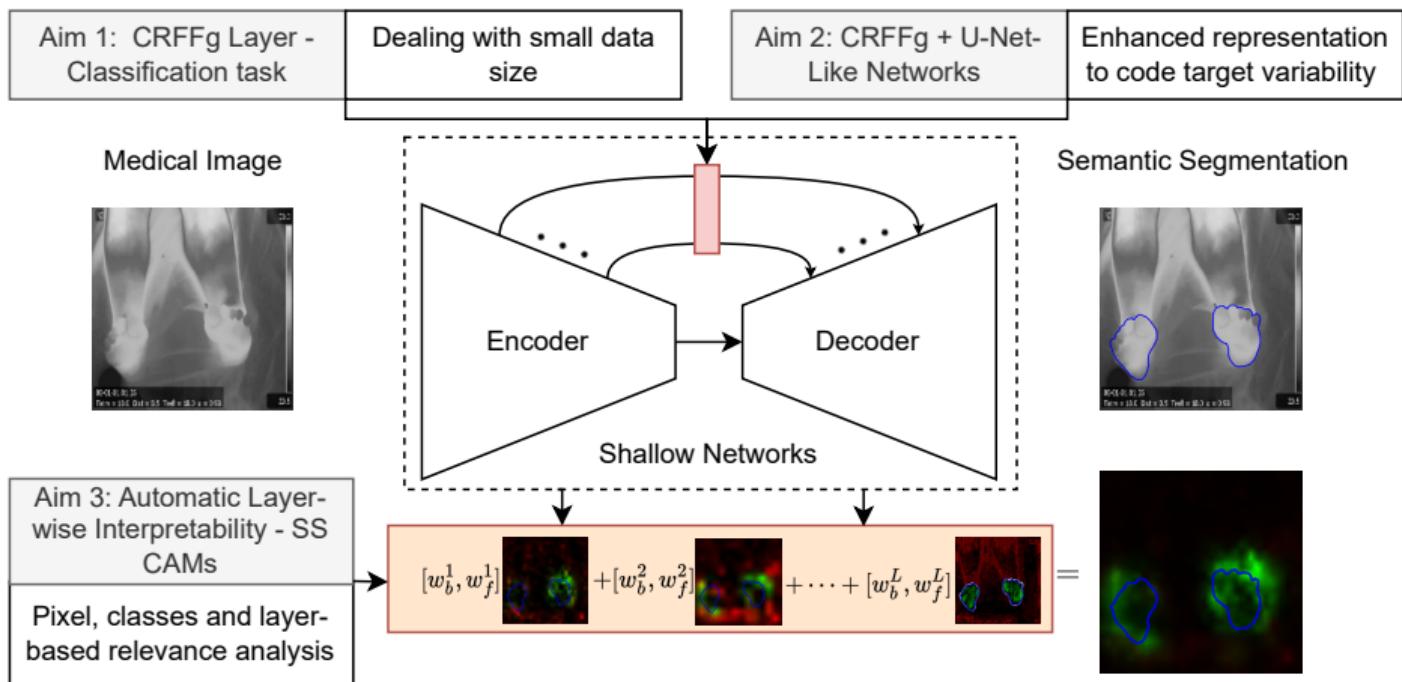
7 Results

8 Conclusions

9 References

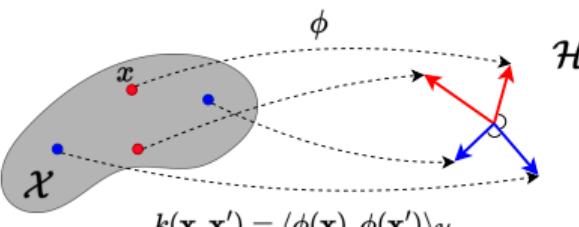


# Methodology





# Convolutional Random Fourier Features Gradient - CRFFg



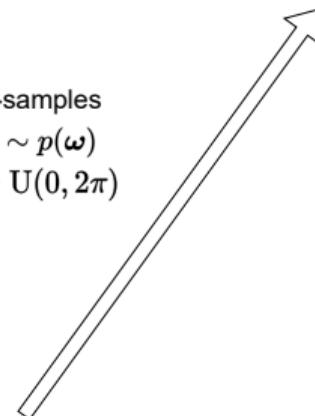
Shift-invariant Kernels

$$k(\mathbf{x} - \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}} \approx \mathbf{z}(\mathbf{x}')^\top \mathbf{z}(\mathbf{x}')$$

Bochner's Theorem

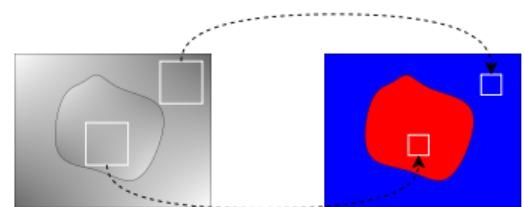
$$k(\mathbf{x} - \mathbf{x}') = \int_{\mathbb{R}^Q} p(\omega) \exp(i\omega^\top (\mathbf{x} - \mathbf{x}')) d\omega = \mathbb{E}_{\omega} \{ \exp(i\omega^\top \mathbf{x}) \exp(-i\omega^\top \mathbf{x}') \}$$

Q-samples  
 $\omega \sim p(\omega)$   
 $b \sim U(0, 2\pi)$



$$\mathbf{z}(\mathbf{x}) = \sqrt{\frac{2}{Q}} [\cos(\omega_1^\top \mathbf{x} + b_1), \dots, \cos(\omega_Q^\top \mathbf{x} + b_Q)]^\top$$

Localities and translation equivariance



$$\mathbf{F}_l = z(\mathbf{F}_{l-1}) = \cos \left( \frac{\mathbf{W}_l}{\Delta_l} \otimes \mathbf{F}_{l-1} + \mathbf{b}_l \right)$$

$\otimes$  stands for convolutional 2-D operation.



# Deep Learning Framework for SS

## ■ Dataset:

$$\{\mathbf{I}_n \in \mathbb{R}^{R \times C}, \mathbf{M}_n \in \{0, 1\}^{R \times C} : n \in N\}$$

$\mathbf{I}_n$  and  $\mathbf{M}_n$  are the  $n$ -th image and target mask, respectively.

## ■ Model:

$$\hat{\mathbf{M}} = (\varphi_L \circ \dots \circ \varphi_1)(\mathbf{I})$$

$\varphi_I$  is the  $I$ -th convolutional layer with parameters  $\{\mathbf{W}_I \in \mathbb{R}^{P_I \times P_I \times D_I} : I \in L\}$ .

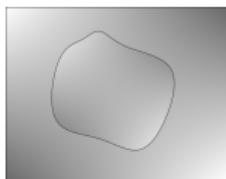
## ■ Dice-based optimization problem:

$$\{\mathbf{W}_I\}_{I=1}^L = \arg \min_{\mathbf{W}_I} \mathbb{E} \left\{ -2 \frac{\mathbf{1}^\top (\mathbf{M}_n \odot \hat{\mathbf{M}}_n) \mathbf{1} + \epsilon}{\mathbf{1}^\top \mathbf{M}_n \mathbf{1} + \mathbf{1}^\top \hat{\mathbf{M}}_n \mathbf{1} + \epsilon} : \forall n \in \{1, 2, \dots, N\} \right\},$$

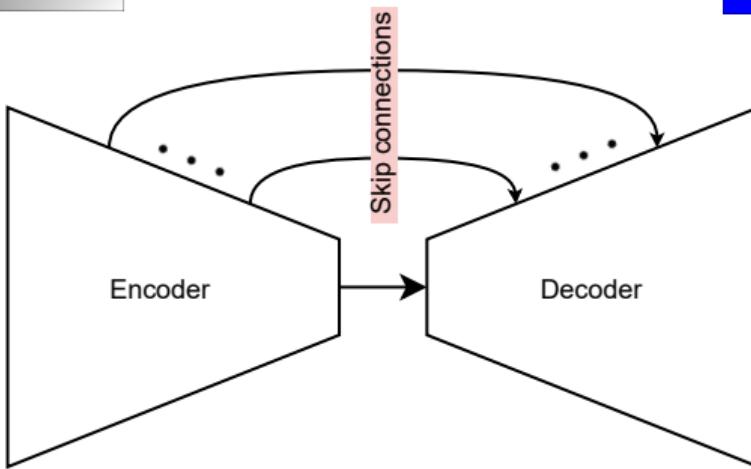
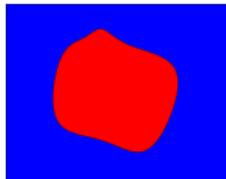
$\odot$  stands for the element-wise product, and  $\epsilon=1$  avoids numerical instability.



# Encoder-Decoder Enhancement using CRFFg



$$\mathbf{F}_l = z(\mathbf{F}_{l-1}) = \cos\left(\frac{\mathbf{W}_l}{\Delta_l} \otimes \mathbf{F}_{l-1} + \mathbf{b}_l\right)$$



Improving characterization at the skip connection through CRFFg



# CAM Methods

## ■ Grad-CAM [Selvaraju et al., 2016]

$$[\beta_I^{cd}]_{i,j} = \frac{1}{H_I W_I} \sum_{n \in H_j, m \in W_I} \frac{\partial y^c}{\partial \mathbf{A}_{nm}^{cd}} \quad \forall \quad i, j$$

## ■ Grad-CAM ++ [Chattopadhyay et al., 2017]

$$[\beta_I^{cd}]_{i,j} = \sum_{n \in H_j, m \in W_I} \alpha_{nm}^{cd} \text{ReLU}\left(\frac{\partial y^c}{\partial \mathbf{A}_{nm}^d}\right) \quad \forall \quad i, j$$

$$\alpha_{nm}^{cd} = \frac{\frac{\partial^2 y^c}{(\partial \mathbf{A}_{nm}^d)^2}}{2 \frac{\partial^2 y^c}{(\partial \mathbf{A}_{nm}^d)^2} + \sum_a \sum_b \mathbf{A}_{ab}^d \frac{\partial^3 y^c}{(\partial \mathbf{A}_{nm}^d)^3}}$$

## ■ Score-CAM [Wang et al., 2019]

$$[\beta_I^{cd}]_{i,j} = \frac{\exp(\xi_d^c)}{\sum_n \exp(\xi_n^c)} \quad \forall \quad i, j$$

$$\xi_d^c = f^c(\mathbf{I} \circ \mathbf{A}^d) - f^c(\mathbf{I})$$

## ■ Layer-CAM [Jiang et al., 2021]

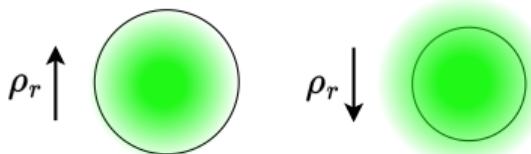
$$[\beta_I^{cd}]_{i,j} = \text{ReLU}\left(\frac{\partial y^c}{\partial \mathbf{A}_{ij}^d}\right)$$



# Layer-Wise Weighted Class Activation Maps

- **CAM-based Cumulative Relevance  $\rho_r$**

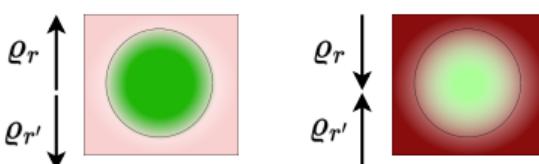
$$\rho_r = \mathbb{E}_I \left\{ \mathbb{E}_n \left\{ \frac{\mathbf{1}^\top (\tilde{\mathbf{M}}_n^r \odot \mathbf{S}_{nl}^r) \mathbf{1}}{\mathbf{1}^\top \tilde{\mathbf{M}}_n^r \mathbf{1}} : \forall n \in N \right\} : \forall I \in L \right\}, \quad \rho_r \in [0, 1]$$



- **Mask-based Cumulative Relevance  $\varrho_r$**

$$\varrho_{rl}' = \mathbb{E}_n \left\{ \frac{\mathbf{1}^\top (\tilde{\mathbf{M}}_n^r \odot \mathbf{S}_{nl}^r) \mathbf{1}}{\mathbf{1}^\top \tilde{\mathbf{M}}_n^r \mathbf{1}} : \forall n \in N \right\}, \quad \varrho_{rl}' \in \mathbb{R}^+$$

$$\varrho_r = \mathbb{E}_I \left\{ \frac{\varrho_{rl}'}{\max_{c \in \{0,1\}} \varrho_{cl}'} : \forall I \in L \right\}, \quad \varrho_r' \in [0, 1]$$



- **CAM-Dice  $D'_r$**

$$D'_r = \mathbb{E}_I \left\{ \mathbb{E}_n \left\{ 2 \frac{\mathbf{1}^\top (\tilde{\mathbf{M}}_n^r \odot \mathbf{S}_{nl}^r) \mathbf{1}}{\mathbf{1}^\top \tilde{\mathbf{M}}_n^r \mathbf{1} + \mathbf{1}^\top \mathbf{S}_{nl}^r \mathbf{1}} : \forall n \in N \right\} : \forall I \in L \right\}, \quad D'_r \in [0, 1]$$





# Outline

- 1 Motivation
- 2 Problem Statement
- 3 Literature Review
- 4 Aims
- 5 Methodology
- 6 Experimental Set-up
- 7 Results
- 8 Conclusions
- 9 References



# Classification: Dataset - FashionMnist



70000 images of 24x24: 60000 for training and 10000 for testing.



# Classification: Method Comparison

- Standard RFF [Rahimi and Recht, 2009, Liu et al., 2021] vs. CRFFg in the projected space varying the output dimension on randomly selected images.
- Method comparison: Full CNN, Full CRFFg, and CNN with an RFF after the flattened layer.

Name Layer	Type	Output Shape	Param #
Input	InputLayer	[(None, 28, 28, 1)]	0
CRFFg1 (Conv01*)	ConvRFF (Conv2D)	(None, 26, 26, 16)	161
MaxPool01	MaxPooling2D	(None, 13, 13, 16)	0
CRFFg2 (Conv02*)	ConvRFF (Conv2D)	(None, 11, 11, 32)	4640
MaxPool02	MaxPooling2D	(None, 5, 5, 32)	0
CRFFg3 (Conv03*)	ConvRFF (Conv2D)	(None, 3, 3, 64)	18486
Flatten	Flatten	(None, 576)	0
Dense( $RFF^+$ )	Dense(RFF)	(None, 32)	18464
Output	Dense	(None, 10)	330



# Classification: Performance measures

Comparison in the mapped space:

$$\epsilon_{n,I} = |\langle \mathbf{z}_{RFF}(\mathbf{x}'_n; R_I), \mathbf{z}_{RFF}(\mathbf{x}'_n; R_I) \rangle - \langle \mathbf{z}_{CRFFg}(\mathbf{x}_n; R_I), \mathbf{z}_{CRFFg}(\mathbf{x}_n; R_I) \rangle|$$

Comparison in classification task:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$



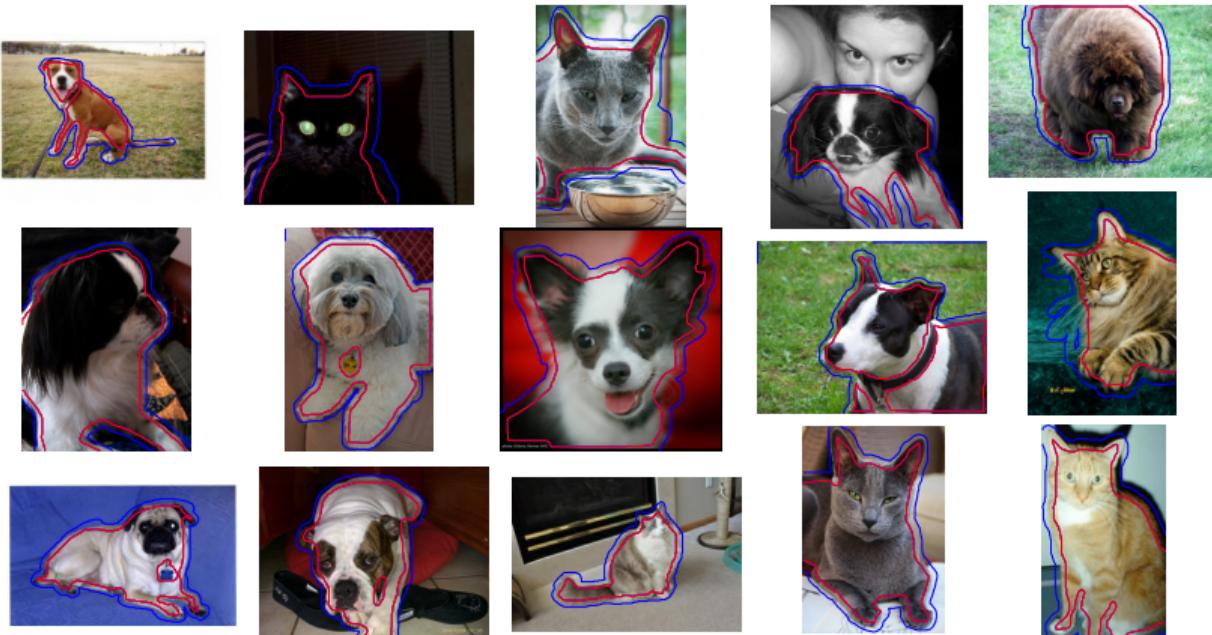
# Segmentation: Dataset - ThermalFeet



166 images, 80% of the samples for training, 10% for validation, and 10% for testing.  
We test with and without data augmentation.



# Segmentation: Dataset - Oxford Iiit Pet



37 different pet categories, each with approximately 200 images, resulting in 3,680 for training and 3,669 for testing. Significant scale, pose, and lighting variations [Parkhi et al., 2012]



# Segmentation: Method Comparison

Table: Variations for each Baseline Model (FCN, ResUNet, U-Net)

Modification \ M	No M	M=1	M=3
Baseline	x		
CRFFg		x	x
Stand Conv		x	x

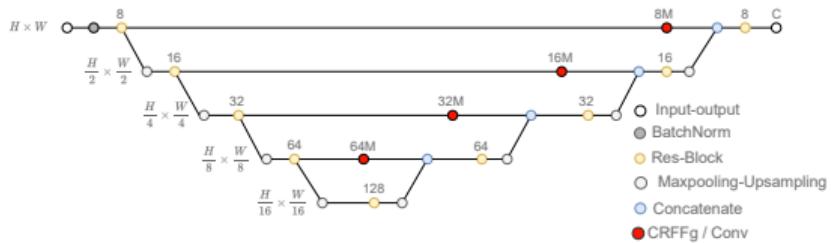


Figure: ResUNet Architecture [Anas et al., 2017]



Figure: FCN Architecture [Long et al., 2014]

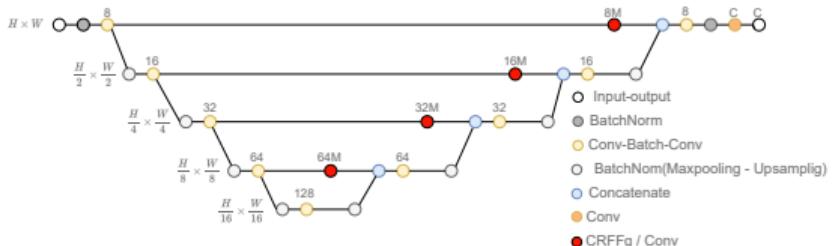
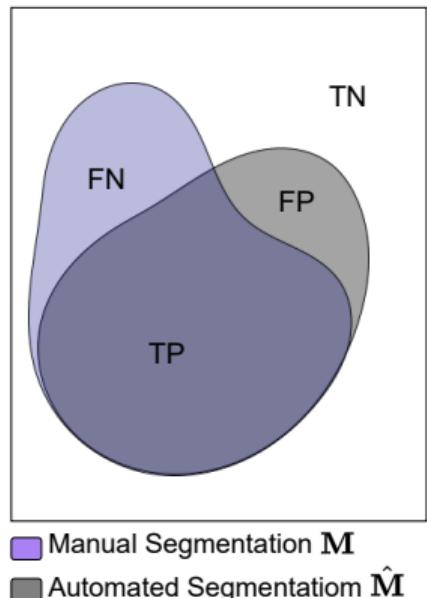


Figure: U-Net Architecture [Ronneberger et al., 2015]



# Segmentation: Performance Measures

- Dice[%] =  $100 \frac{2|M \cap \hat{M}|}{|M| + |\hat{M}|} = 100 \frac{2TP}{2TP + FP + FN}$
- Jaccard[%] =  $100 \frac{|M \cap \hat{M}|}{|M \cup \hat{M}|} = 100 \frac{TP}{FN + FP + TP}$
- Sensitivity[%] =  $100 \frac{|M \cap \hat{M}|}{|M \cap \hat{M}| + |M \cap \neg \hat{M}|} = 100 \frac{TP}{TP + FN}$
- Specificity[%] =  $100 \frac{|\neg M \cap \neg \hat{M}|}{|\neg M \cap \hat{M}| + |\neg M \cap \hat{M}|} = 100 \frac{TN}{TN + FP}$





# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

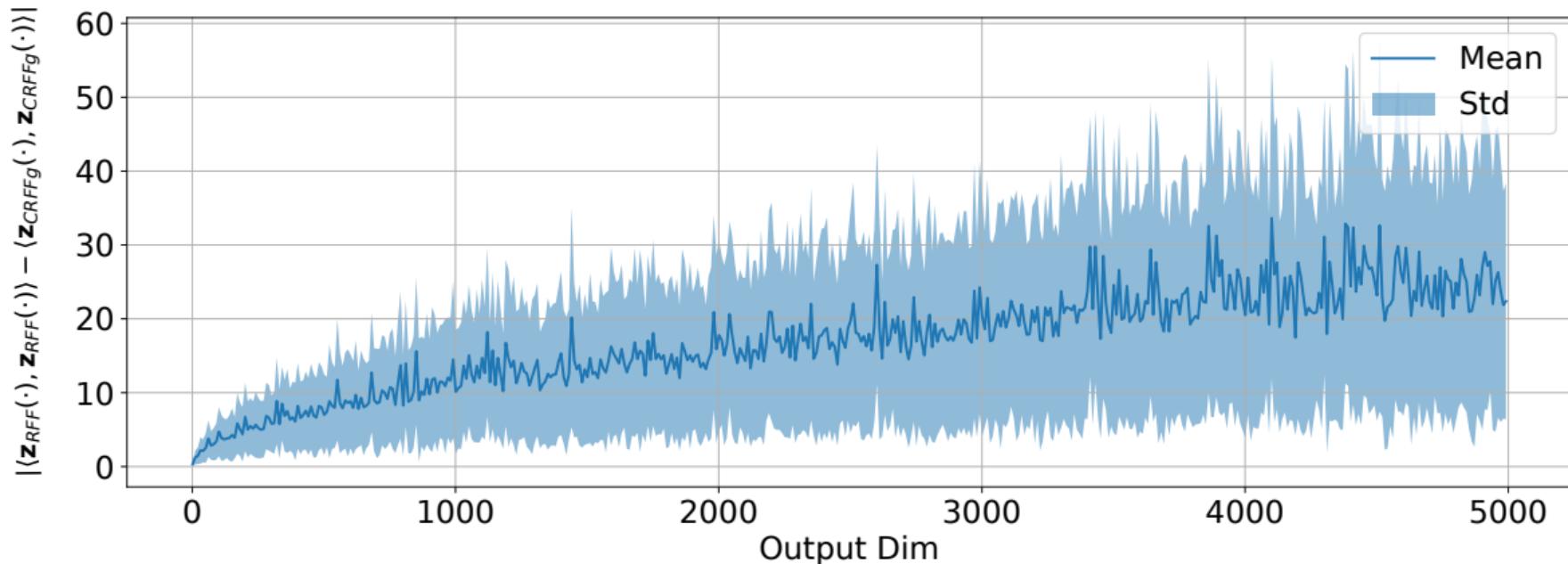
7 Results

8 Conclusions

9 References



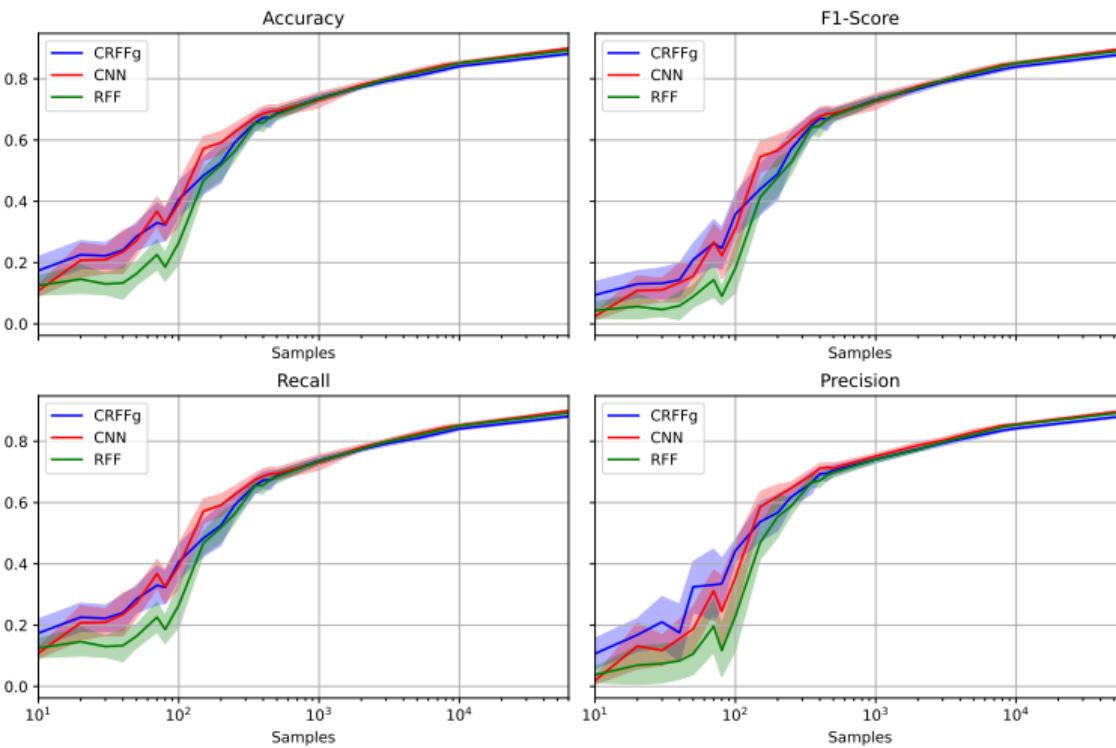
# Classification: RFF vs CRFFg representation



The difference remains relatively low compared to the output dimension



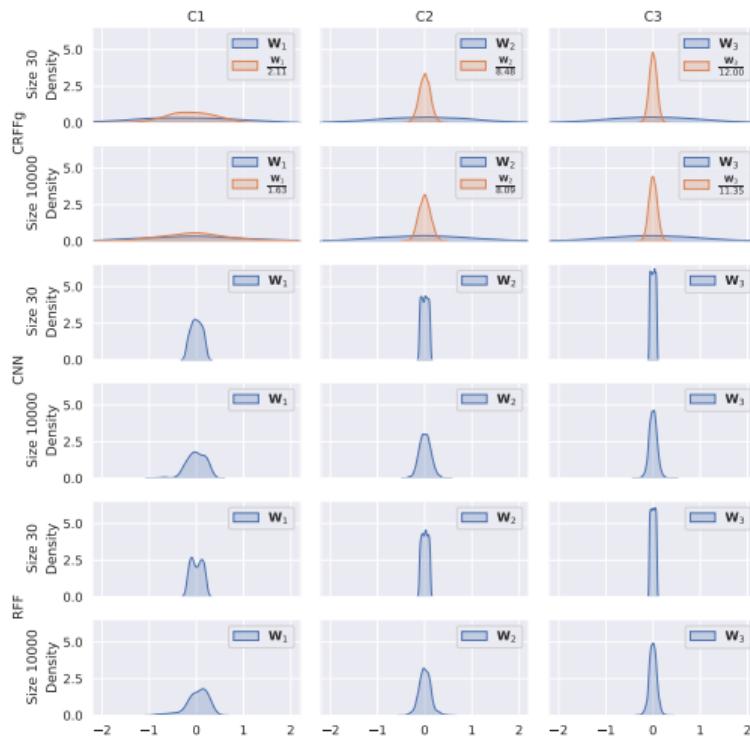
# Classification: performance vs training size



CRFFg can capture both kernel methods' generalization capability under low sample sizes and the local and equivariant characterization of CNNs.



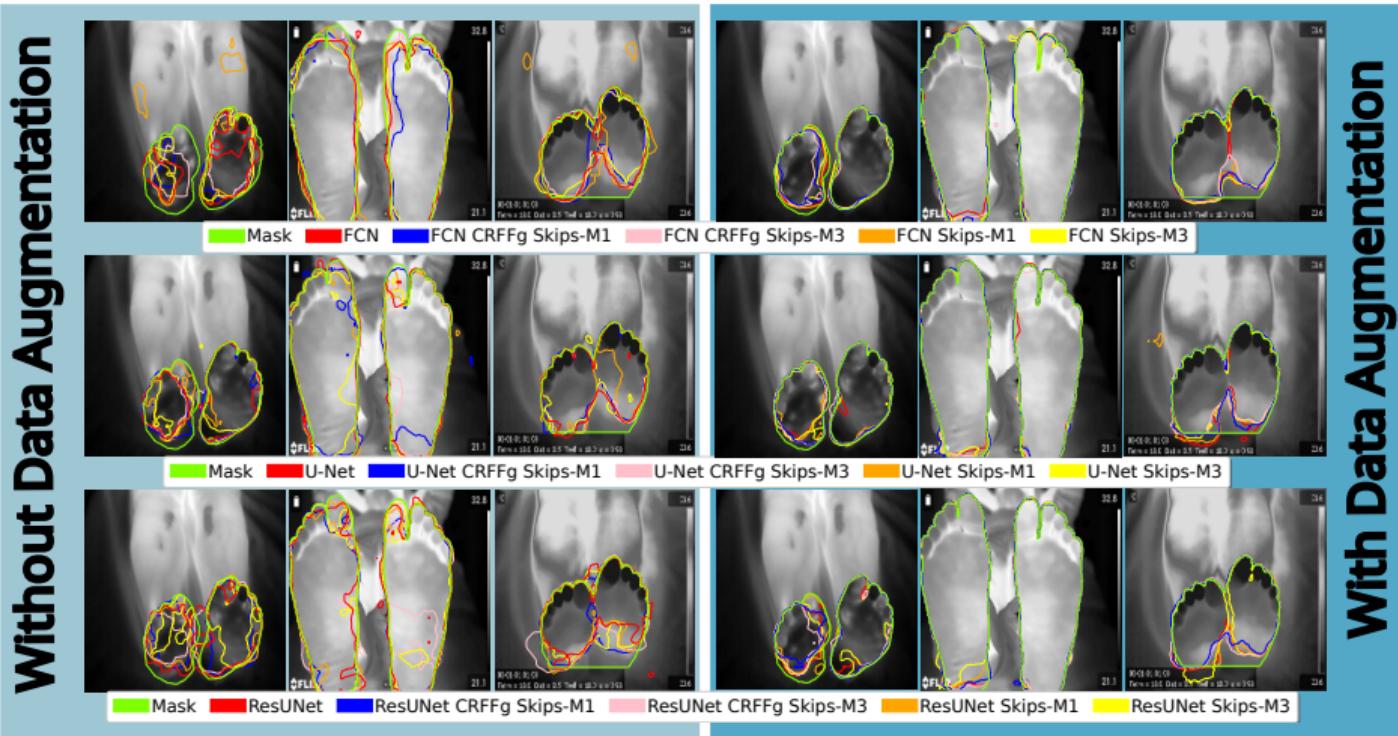
# Classification: Weights Distributions



Weight behavior is significantly influenced by the value of  $\sigma$



# Segmentation: Visual Inspection Results I



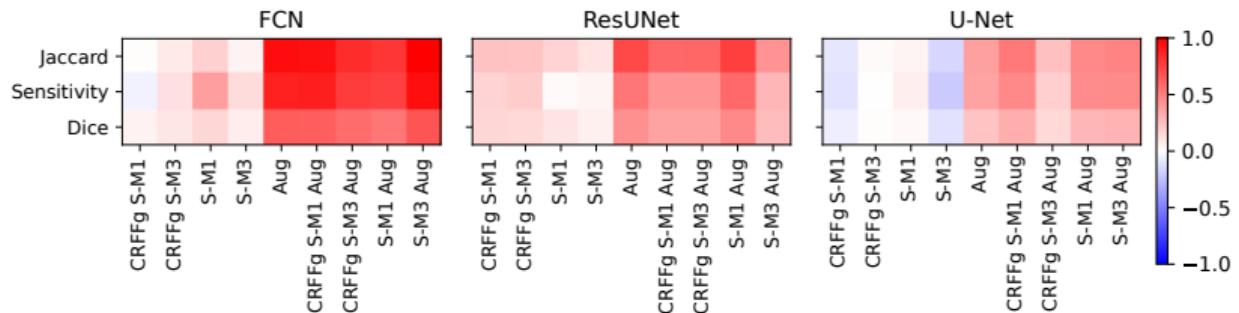
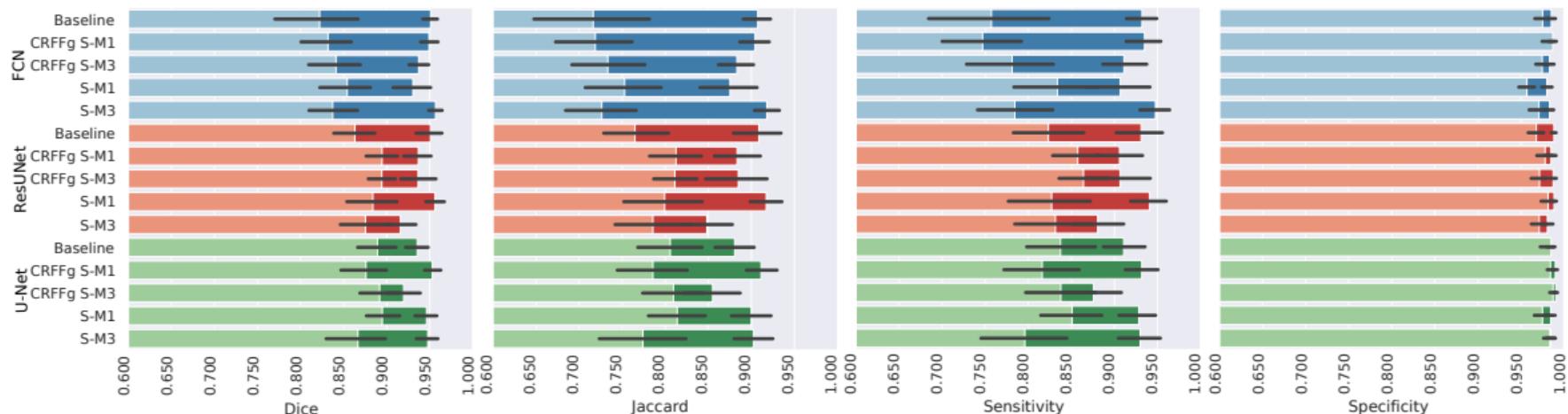


# Segmentation: Visual Inspection Results II



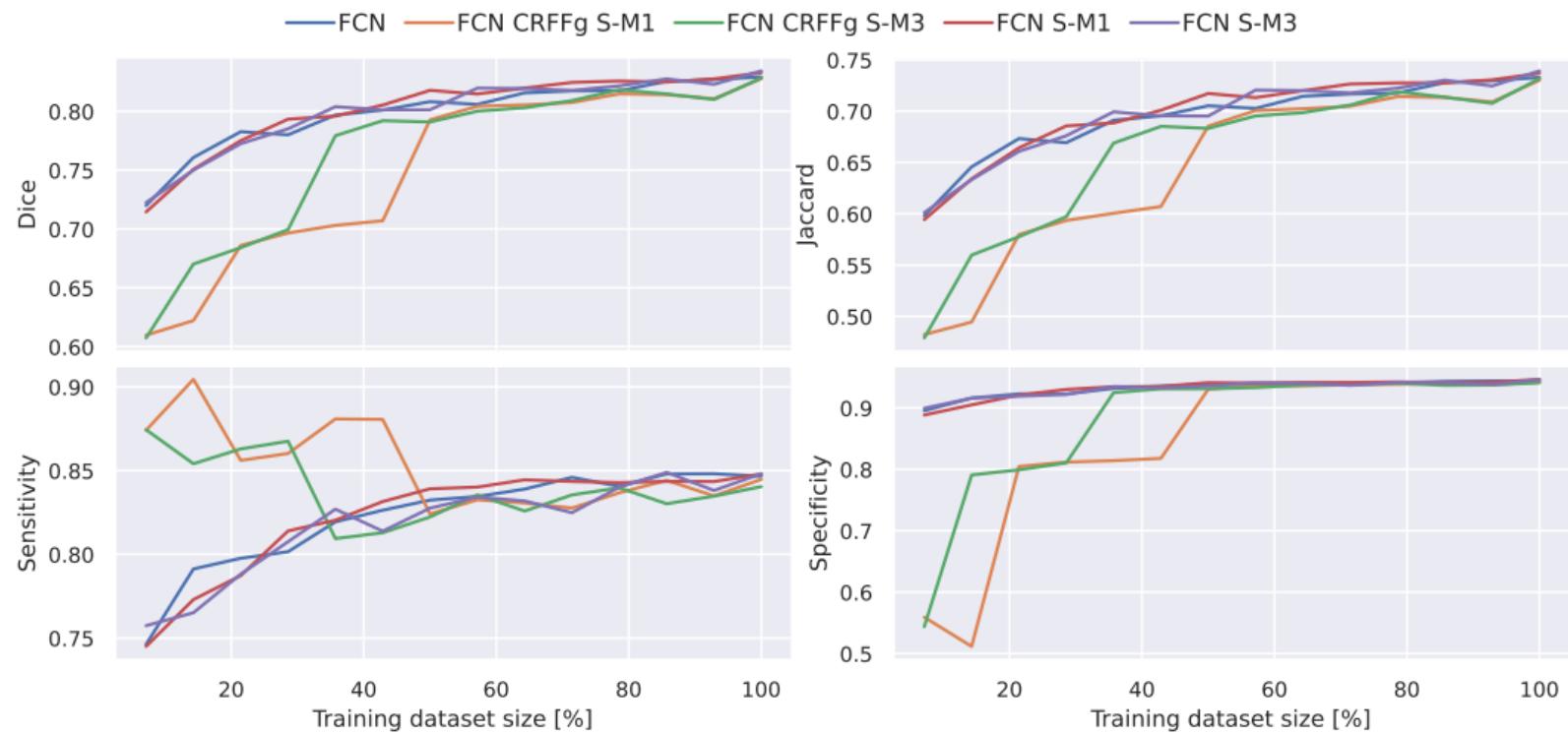


# Quantitative Semantic Segmentation Results - ThermalFeet



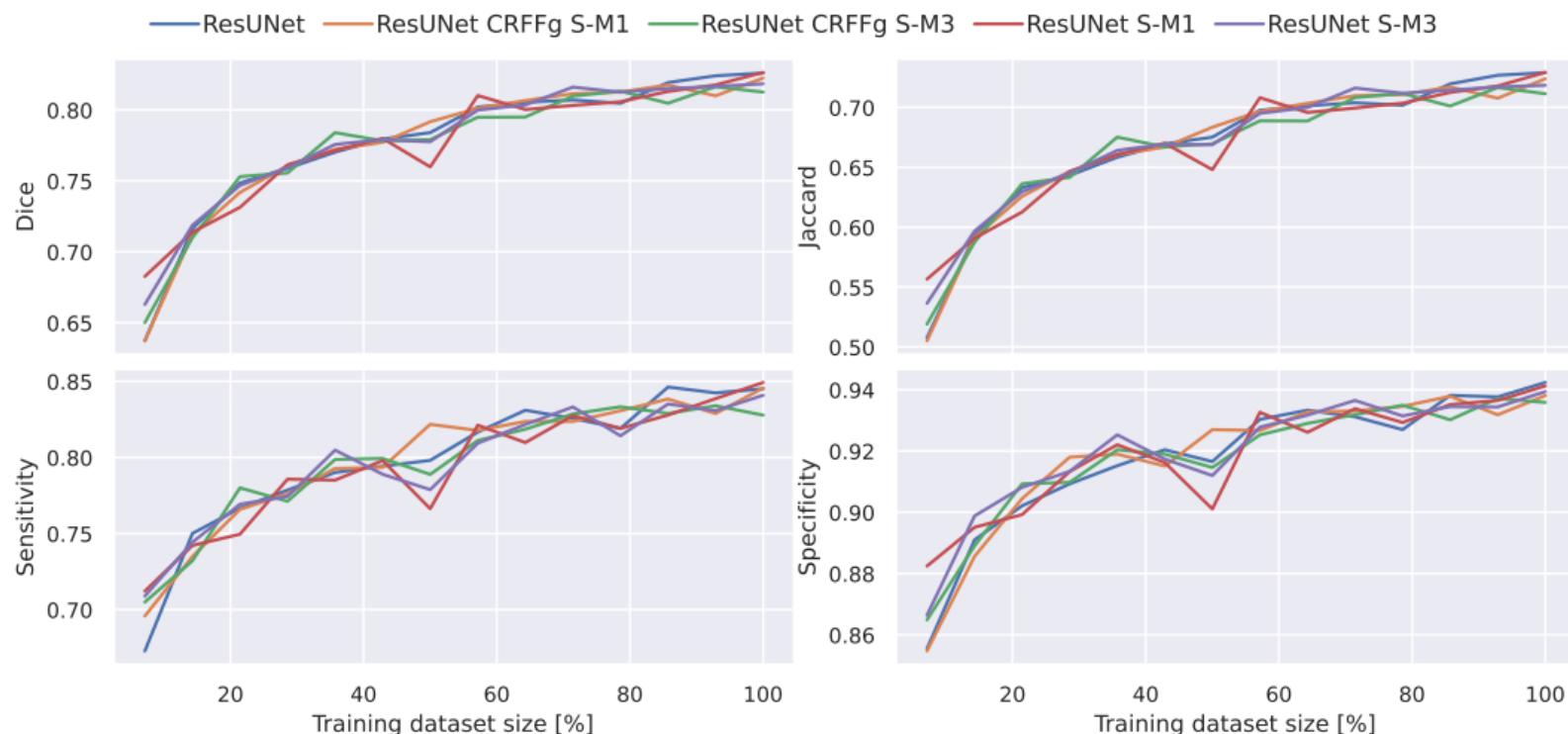


# Performance vs Training dataset size - Oxford Iiit Pet I



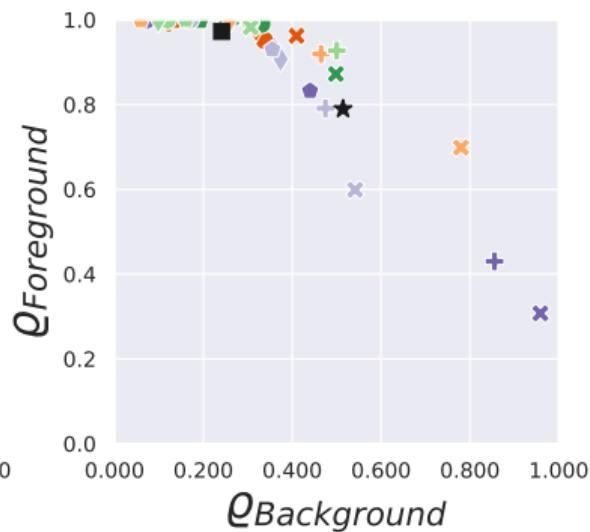
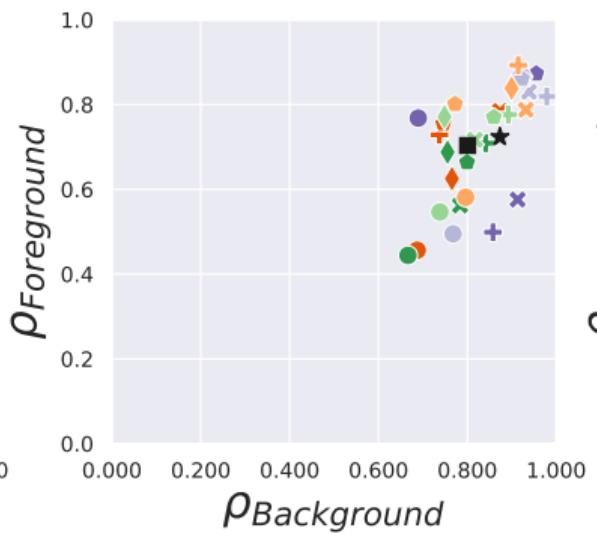
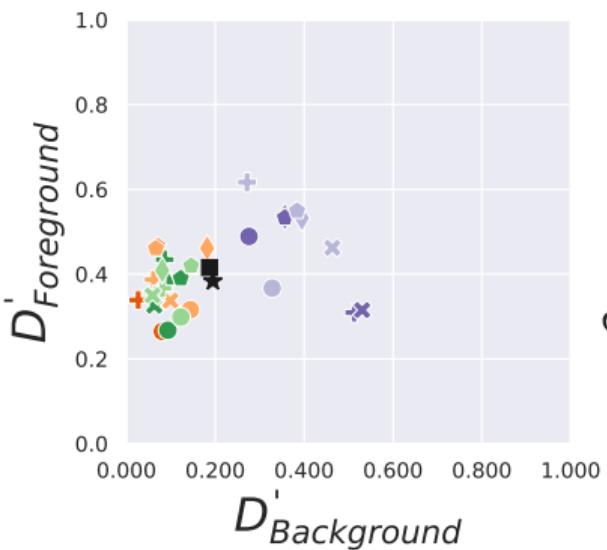
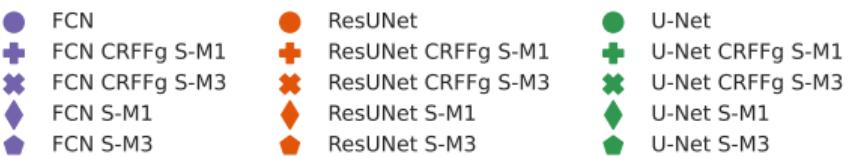


# Performance vs Training dataset size - Oxford Iiit Pet II





# Interpretability Results - ThermalFeet I





# Interpretability Results - ThermalFeet II

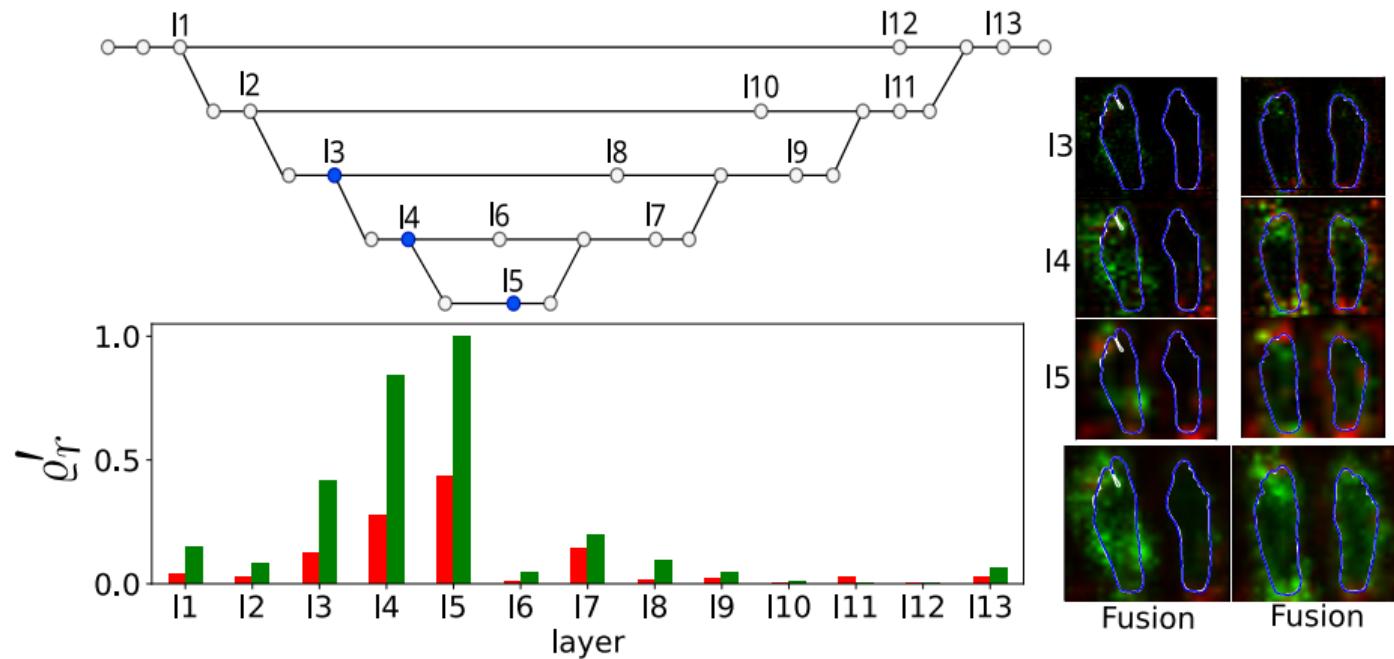


Figure: ResUNet CRFFg S-M3 without data Augmentation



# Integration into a software tool for medical use I



Thermographic Camera



Computational System

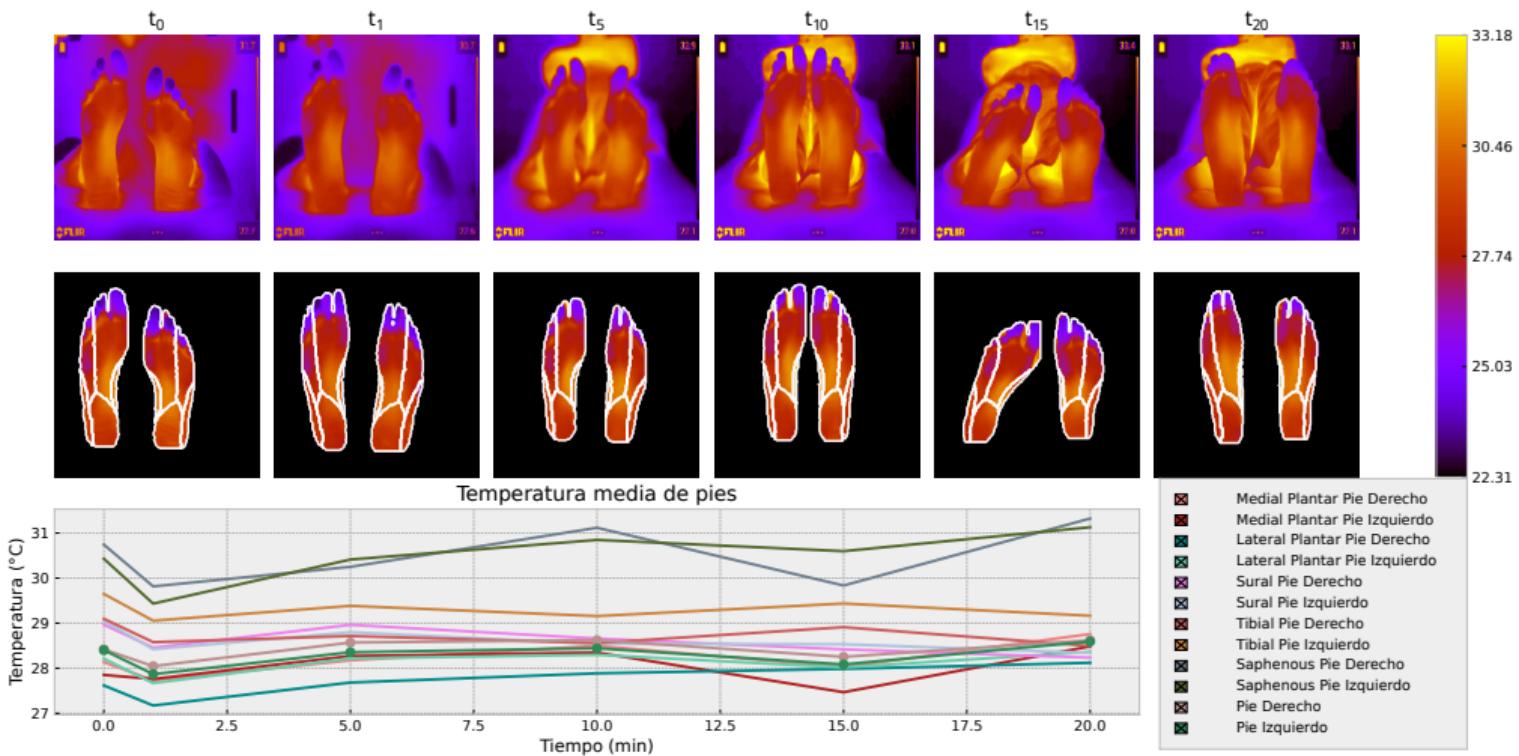


Interface

Monitoring system for temperature changes in feet soles able to work in obstetrics environment  
(easy to use for the medical staff)



# Integration into a software tool for medical use II





# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

7 Results

8 Conclusions

9 References



# Conclusions

- We propose an extension of the **Random Fourier Features** tailored for **spatial data**, introducing **CRFFg**—a data-driven approach that leverages gradient descent.
- Our innovative extension effectively **enhances feature representation** within the skip connection of encoder-decoder models, particularly for semantic segmentation tasks.
- We tested 15 model variations of three well-known deep learning architectures for automatic feet semantic segmentation on **thermal images** that exhibit small **sample size and high variability** of regions of interest.
- We have introduced **quantitative measures** for enhancing **interpretability** in **semantic segmentation** models used in the medical field.
- Our approach **quantifies the relevance location** of specific regions, the sensibility across multiple regions of interest, and their homogeneity.



# Future Work

- **Analyzing the spectral representation** of the CRFFg to gain insights into underlying patterns, leading to a deeper understanding and potential improvements [Zhang et al., 2020].
- Incorporating **Bayesian approximation** techniques to quantify uncertainties, model relationships between variables more accurately, and potentially uncover new strategies related to our CRFFg layer [Miller and Reich, 2022].
- Employing **regularization techniques** to mitigate overfitting using the proposed interpretability measures [Chang et al., 2020, Lin et al., 2021].
- Exploring alternative mappings for Random Fourier Features [Sutherland and Schneider, 2015] and the training of the scale  $\sigma$  parameter, addressing class imbalance with different loss functions [Yeung et al., 2022], and considering transformers instead of convolutions for better context capture [Azad et al., 2023].



# Academic Discussion I

- Aguirre-Arango, J.C.; Álvarez-Meza, A.M.; Castellanos-Dominguez, G. Feet Segmentation for Regional Analgesia Monitoring Using Convolutional RFF and Layer-Wise Weighted CAM Interpretability. *Computation* 2023, 11, 113. <https://doi.org/10.3390/computation11060113> **Q2/A2**
- Mejia-Zuluaga, Rafael, Juan Carlos Aguirre-Arango, Diego Collazos-Huertas, Jessica Daza-Castillo, Néstor Valencia-Marulanda, Mauricio Calderón-Marulanda, Óscar Aguirre-Ospina, Andrés Alvarez-Meza, and Germán Castellanos-Dominguez. "Deep Learning Semantic Segmentation of Feet Using Infrared Thermal Images." In *Advances in Artificial Intelligence—IBERAMIA 2022: 17th Ibero-American Conference on AI*, Cartagena de Indias, Colombia, November 23–25, 2022, Proceedings, pp. 342–352. Cham: Springer International Publishing, 2023. **International Conference**
- A.D. Tobar, J.C. Aguirre, D.A. Cardenas-Pena, A.M. Alvarez-Meza, and C.G. Castellanos-Dominguez, "Hippocampus Segmentation using Patch-based Representation and ROC Label Enhancement," *Engineering Letters*, vol. 31, no. 2, pp504-510, 2023 **B**



# Academic Discussion II

Title	Github repository
Image segmentation library	<a href="https://github.com/UN-GCPDS/python-gcpds.image_segmentation">https://github.com/UN-GCPDS/python-gcpds.image_segmentation</a>
CRFFg layer	<a href="https://github.com/aguirrejuan/ConvRFF">https://github.com/aguirrejuan/ConvRFF</a>
Monitoring tool	<a href="https://github.com/UN-GCPDS/FEET-GUI">https://github.com/UN-GCPDS/FEET-GUI</a>



# Acknowledgments

- “Herramienta de apoyo a la predicción de los efectos de anestésicos locales vía neuroaxial epidural a partir de termografía por infrarrojo” (Code 111984468021) funded by MINCIENCIAS
- “Sistema prototipo de visión por computador utilizando aprendizaje profundo como soporte al monitoreo de zonas urbanas desde unidades aéreas no tripuladas” (Hermes Code 55261), funded by Universidad Nacional de Colombia.

# Thank you!

Juan Carlos Aguirre Arango  
[jucaguirrear@unal.edu.co](mailto:jucaguirrear@unal.edu.co)



# Outline

1 Motivation

2 Problem Statement

3 Literature Review

4 Aims

5 Methodology

6 Experimental Set-up

7 Results

8 Conclusions

9 References



# References I



Alokasi, H. and Ahmad, M. B. (2022).

Deep learning-based frameworks for semantic segmentation of road scenes.

*Electronics*, 11(12).



Anand, T., Sinha, S., Mandal, M., Chamola, V., and Yu, F. R. (2021).

Agrisegnet: Deep aerial semantic segmentation framework for iot-assisted precision agriculture.

*IEEE Sensors Journal*, 21(16):17581–17590.



Anas, E. M. A., Nouranian, S., Mahdavi, S. S., Spadlinger, I., Morris, W. J., Salcudean, S. E., Mousavi, P., and Abolmaesumi, P. (2017).

Clinical target-volume delineation in prostate brachytherapy using residual neural networks.

In Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D. L., and Duchesne, S., editors, *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, pages 365–373, Cham. Springer International Publishing.



Arteaga-Marrero, N., Hernández, A., Villa, E., González-Pérez, S., Luque, C., and Ruiz-Alzola, J. (2021).

Segmentation approaches for diabetic foot disorders.

*Sensors*, 21(3):934.



Asghar, S., Bjerregaard, L. S., Lundstrøm, L. H., Lund, J., Jenstrup, M. T., and Lange, K. H. W. (2014).

Distal infrared thermography and skin temperature after ultrasound-guided interscalene brachial plexus block: a prospective observational study.

*European journal of anaesthesiology*, 31(11):626—634.



Azad, R., Kazerouni, A., Heidari, M., Aghdam, E. K., Molaei, A., Jia, Y., Jose, A., Roy, R., and Merhof, D. (2023).

Advances in medical image analysis with vision transformers: A comprehensive review.

*arXiv preprint arXiv:2301.03505*.



# References II



Bron, E. E., Smits, M., van der Flier, W. M., Vrenken, H., Barkhof, F., Scheltens, P., Papma, J. M., Steketee, R. M., Méndez Orellana, C., Meijboom, R., Pinto, M., Meireles, J. R., Garrett, C., Bastos-Leite, A. J., Abdulkadir, A., Ronneberger, O., Amoroso, N., Bellotti, R., Cárdenas-Peña, D., Álvarez Meza, A. M., Dolph, C. V., Iftekharuddin, K. M., Eskildsen, S. F., Coupé, P., Fonov, V. S., Franke, K., Gaser, C., Ledig, C., Guerrero, R., Tong, T., Gray, K. R., Moradi, E., Tohka, J., Routier, A., Durrleman, S., Sarica, A., Di Fatta, G., Sensi, F., Chincarini, A., Smith, G. M., Stoyanov, Z. V., Sørensen, L., Nielsen, M., Tangaro, S., Inglesi, P., Wachinger, C., Reuter, M., van Swieten, J. C., Niessen, W. J., and Klein, S. (2015). Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural mri: The caddementia challenge. *NeuroImage*, 111:562–579.



Burel, S., Evans, A., and Anghel, L. (2022).

Improving dnn fault tolerance in semantic segmentation applications.

2022 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), pages 1–6.



Cakir, S., Gauß, M., Häppeler, K., Ounajjar, Y., Heinle, F., and Marchthaler, R. (2022).

Semantic segmentation for autonomous driving: Model evaluation, dataset generation, perspective comparison, and real-time capability.



Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N. C., and Sardanelli, F. (2021).

Ai applications to medical images: From machine learning to deep learning.

*Physica Medica*, 83:9–24.



Chang, Y.-T., Wang, Q., Hung, W.-C., Piramuthu, R., Tsai, Y.-H., and Yang, M.-H. (2020).

Mixup-cam: Weakly-supervised semantic segmentation via uncertainty regularization.



Chattopadhyay, A., Sarkar, A., Howlader, P., and Balasubramanian, V. N. (2017).

Grad-cam++: Improved visual explanations for deep convolutional networks.

*Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, 2018-January:839–847.



# References III



Haren, F., Kadic, L., and Driessen, J. (2013).

Skin temperature measured by infrared thermography after ultrasound-guided blockade of the sciatic nerve.

*Acta anaesthesiologica Scandinavica*, 57.



Jiang, P. T., Zhang, C. B., Hou, Q., Cheng, M. M., and Wei, Y. (2021).

Layercam: Exploring hierarchical class activation maps for localization.

*IEEE Transactions on Image Processing*, 30:5875–5888.



Jimenez, D. A., García, H. F., Álvarez, A. M., and Orozco, Á. A. (2018).

3d probabilistic morphable models for brain tumor segmentation.

In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 22nd Iberoamerican Congress, CIARP 2017, Valparaíso, Chile, November 7–10, 2017, Proceedings* 22, pages 314–322. Springer.



Jimenez-Castaño, C. A., Álvarez Meza, A. M., Aguirre-Ospina, O. D., Cárdenas-Peña, D. A., and Álvaro Angel Orozco-Gutiérrez (2021).

Random fourier features-based deep learning improvement with class activation interpretability for nerve structure segmentation.

*Sensors* 2021, Vol. 21, Page 7741, 21:7741.



Khan, M. Z., Gajendran, M. K., Lee, Y., and Khan, M. A. (2021).

Deep neural architectures for medical image semantic segmentation: Review.

*IEEE ACCESS*, 9:83002–83024.



Kriti, Virmani, J., and Agarwal, R. (2022).

A characterization approach for the review of cad systems designed for breast tumor classification using b-mode ultrasound images.

*Archives of Computational Methods in Engineering*, pages 1–39.



# References IV



Li, J., Zhu, G., Hua, C., Feng, M., Li, P., Lu, X., Song, J., Shen, P., Xu, X., Mei, L., et al. (2021).

A systematic collection of medical image datasets for deep learning.  
*arXiv preprint arXiv:2106.12864.*



Lilay, M. Y. and Taye, G. D. (2023).

Semantic segmentation model for land cover classification from satellite images in gambella national park, ethiopia.  
*SN Applied Sciences*, 5(3):76.



Lin, D., Li, Y., Prasad, S., Nwe, T. L., Dong, S., and Oo, Z. M. (2021).

Cam-guided u-net with adversarial regularization for defect segmentation.  
In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1054–1058.



Lin, J., Guo, X., Shao, J., Jiang, C., Zhu, Y., and Zhu, S.-C. (2016).

A virtual reality platform for dynamic human-scene interaction.

In *SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, SA '16, New York, NY, USA. Association for Computing Machinery.



Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020).

Explainable ai: A review of machine learning interpretability methods.  
*Entropy*, 23(1):18.



Liu, F., Huang, X., Chen, Y., and Suykens, J. A. K. (2021).

Random features for kernel approximation: A survey on algorithms, theory, and beyond.



# References V



Long, J., Shelhamer, E., and Darrell, T. (2014).

Fully convolutional networks for semantic segmentation.

*CoRR*, abs/1411.4038.



Luo, Z., Yang, W., Yuan, Y., Gou, R., and Li, X. (2023).

Semantic segmentation of agricultural images: A survey.

*Information Processing in Agriculture*.



Melesse, A. S., Bayable, S. D., Simegn, G. D., Ashebir, Y. G., Ayenew, N. T., and Fetene, M. B. (2022).

Survey on knowledge, attitude and practice of labor analgesia among health care providers at debre markos comprehensive specialized hospital, ethiopia 2021. a cross-sectional study.

*ANNALS OF MEDICINE AND SURGERY*, 74.



Miller, M. J. and Reich, B. J. (2022).

Bayesian spatial modeling using random fourier frequencies.

*Spatial Statistics*, 48:100598.



Parkhi, O. M., Vedaldi, A., Zisserman, A., and Jawahar, C. V. (2012).

Cats and dogs.

In *IEEE Conference on Computer Vision and Pattern Recognition*.



Qureshi, I., Yan, J., Abbas, Q., Shaheed, K., Riaz, A. B., Wahid, A., Khan, M. W. J., and Szczuko, P. (2023).

Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends.

*Information Fusion*, 90:316–352.



# References VI



Rahimi, A. and Recht, B. (2009).

Random features for large-scale kernel machines.



Rizzoli, G., Barbato, F., and Zanuttigh, P. (2022).

Multimodal semantic segmentation in autonomous driving: A review of current approaches and future perspectives.  
*Technologies*, 10(4).



Ronneberger, O., Fischer, P., and Brox, T. (2015).

U-net: Convolutional networks for biomedical image segmentation.

*CoRR*, abs/1505.04597.



Salahuddin, Z., Woodruff, H. C., Chatterjee, A., and Lambin, P. (2022).

Transparency of deep neural networks for medical image analysis: A review of interpretability methods.  
*Computers in biology and medicine*, 140:105111.



Sarker, I. H. (2021).

Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions.

*SN Computer Science*, 2(6):420.



Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2016).

Grad-cam: Visual explanations from deep networks via gradient-based localization.

*International Journal of Computer Vision*, 128:336–359.



Soomro, T. A., Zheng, L., Afifi, A. J., Ali, A., Soomro, S., Yin, M., and Gao, J. (2023).

Image segmentation for mr brain tumor detection using machine learning: A review.

*IEEE Reviews in Biomedical Engineering*, 16:70–90.



# References VII

-  Sutherland, D. J. and Schneider, J. (2015).  
On the error of random fourier features.
-  Tiator, M., Kerkmann, A. M., Geiger, C., and Grimm, P. (2020).  
Using semantic segmentation to assist the creation of interactive vr applications.  
In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 1–9.
-  Tsai, J., Chang, C.-C., and Li, T. (2023).  
Autonomous driving control based on the technique of semantic segmentation.  
*Sensors*, 23(2).
-  Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2019).  
Score-cam: Score-weighted visual explanations for convolutional neural networks.
-  Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., and Nandi, A. K. (2022).  
Medical image segmentation using deep learning: A survey.  
*IET Image Processing*, 16(5):1243–1267.
-  Wang, X., Wang, L., Zhong, X., Bai, C., Huang, X., Zhao, R., and Xia, M. (2021).  
Pai-net: A modified u-net of reducing semantic gap for surgical instrument segmentation.  
*IET Image Processing*, 15:2959–2969.
-  Wieland, M., Martinis, S., Kiefl, R., and Gstaiger, V. (2023).  
Semantic segmentation of water bodies in very high-resolution satellite and aerial images.  
*Remote Sensing of Environment*, 287:113452.



# References VIII

-  Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L. R., Summers, R. M., Rubin, D. L., and Lungren, M. P. (2020). Preparing medical imaging data for machine learning. *Radiology*, 295(1):4–15.
-  Yeung, M., Sala, E., Schönlieb, C.-B., and Rundo, L. (2022). Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 95:102026.
-  Zhang, X., Xu, J., Yang, J., Chen, L., Zhou, H., Liu, X., Li, H., Lin, T., and Ying, Y. (2020). Understanding the learning mechanism of convolutional neural networks in spectral analysis. *Analytica Chimica Acta*, 1119:41–51.
-  Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2020). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation.



# Segmentation: Context information + Localization

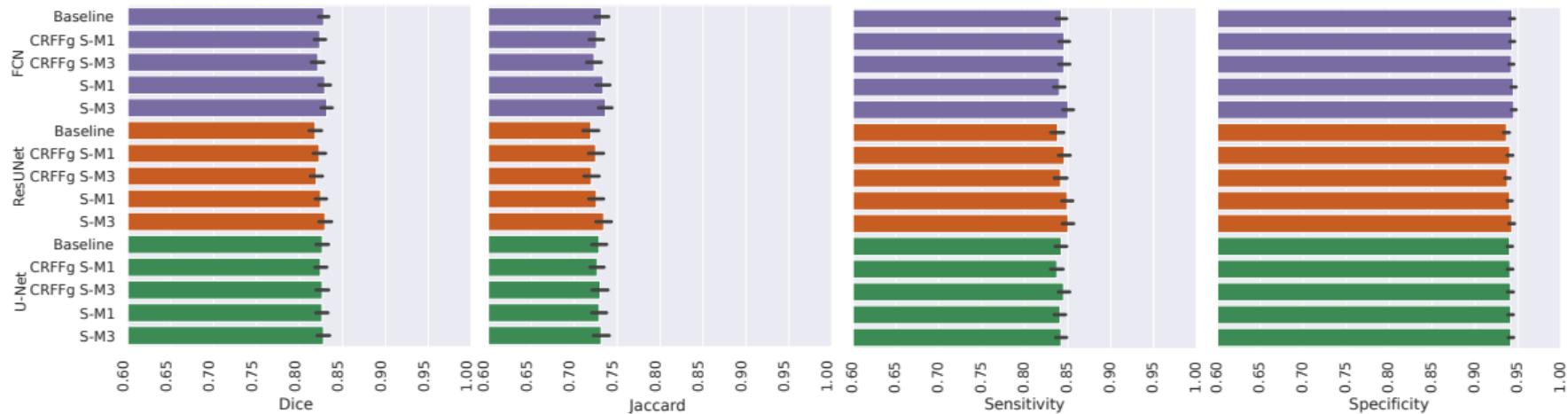
Segmentation: Context information + Localization



Transformers have an infinite receptive field (High context).  
Convolution has a local receptive field. Multiple layers increase the receptive field.



# Quantitative Semantic Segmentation Results - Oxford Pet IIIt





# Interpretability Results - ThermalFeet I

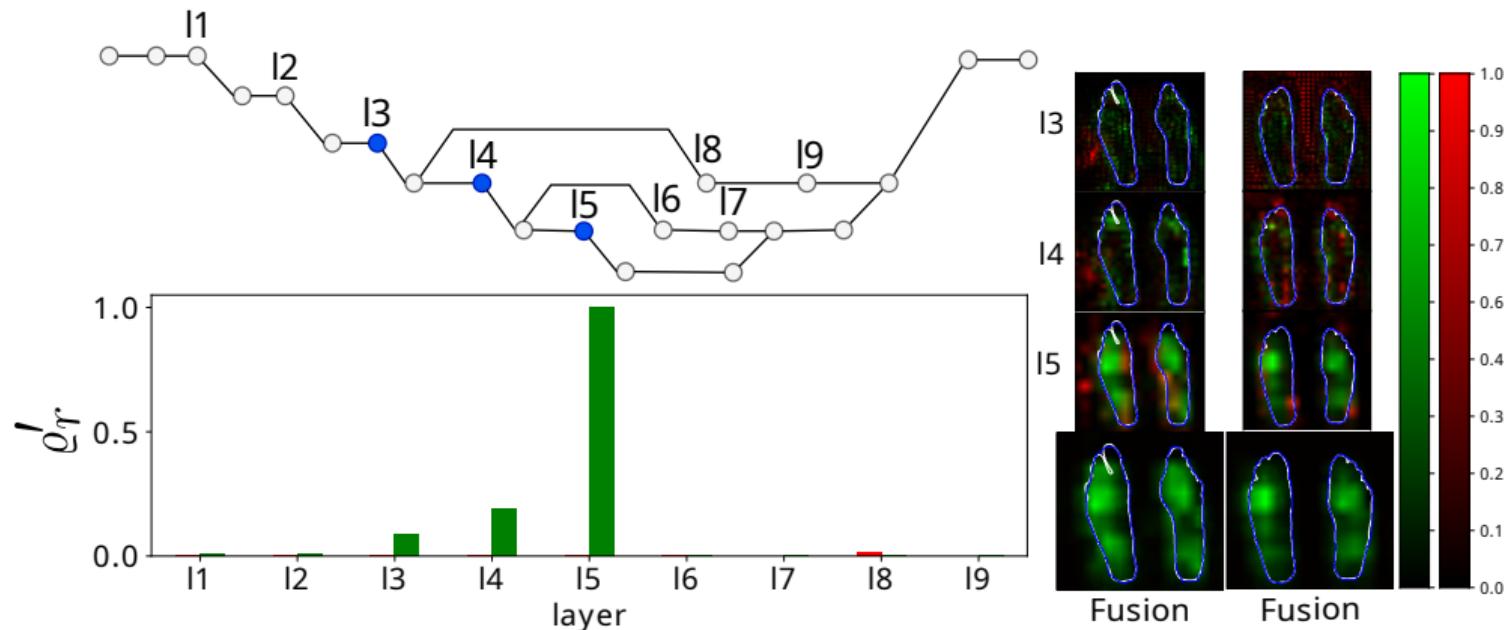


Figure: FCN CRFFg S-M3 without data Augmentation



# Interpretability Results - ThermalFeet II

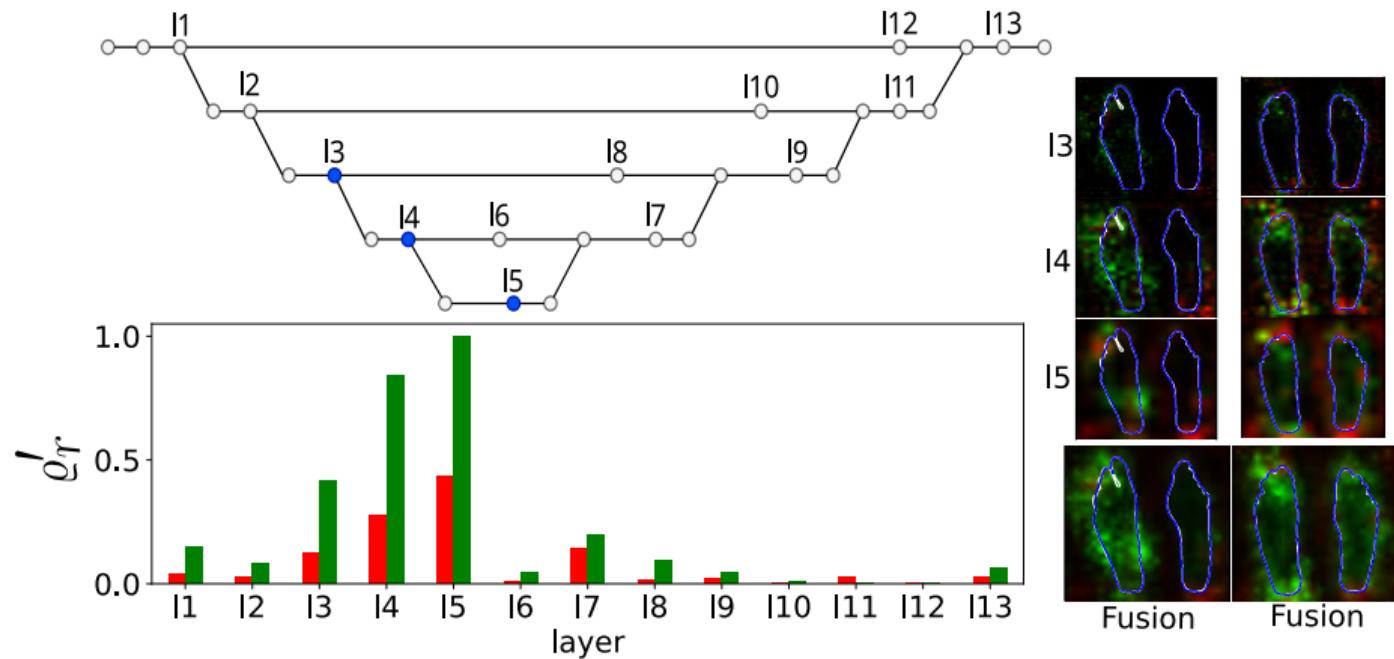


Figure: ResUNet CRFFg S-M3 without data Augmentation



# Interpretability Results - ThermalFeet III

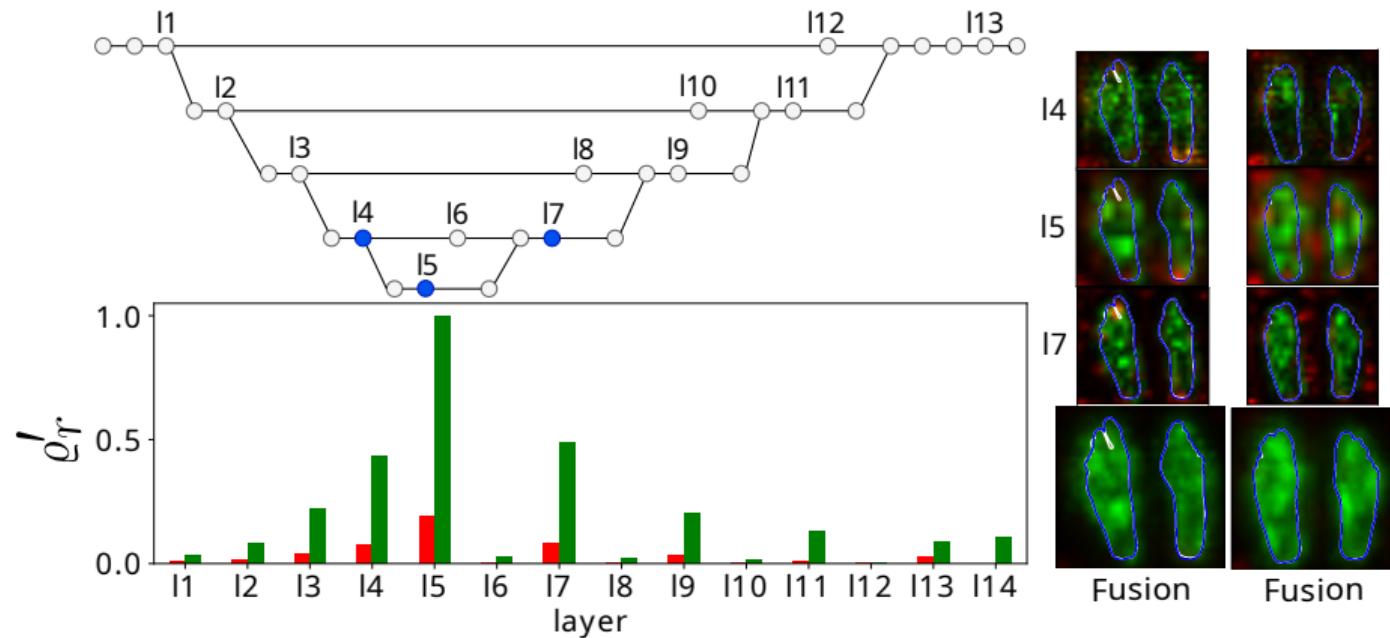
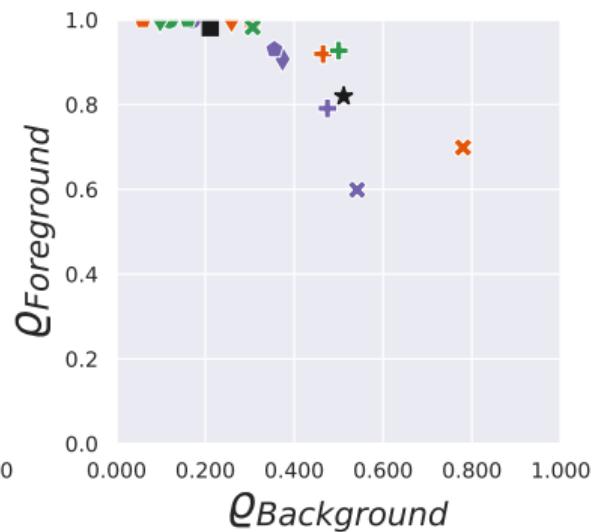
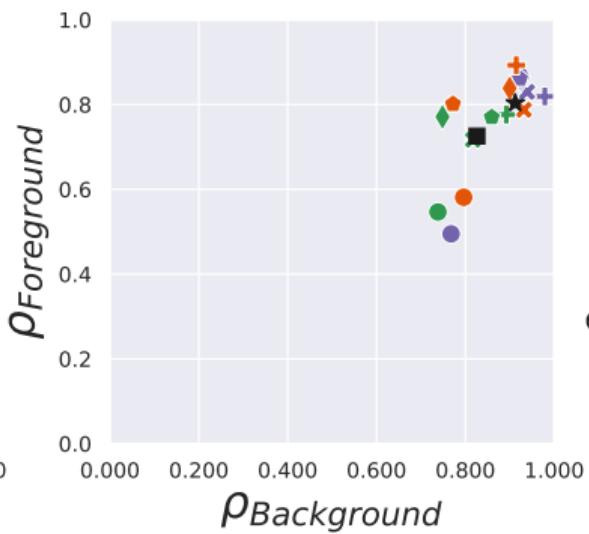
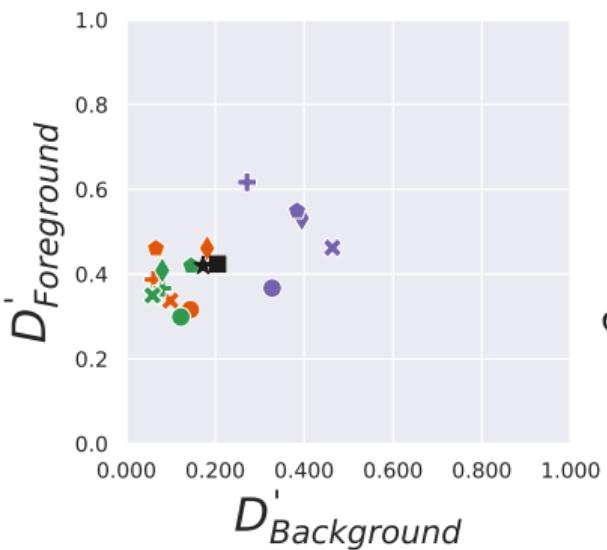
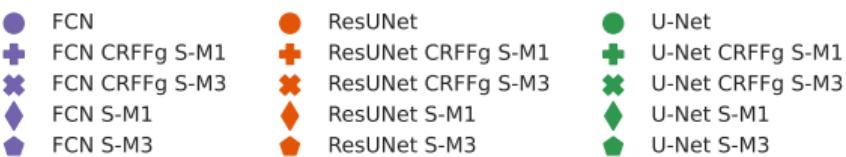


Figure: U-Net CRFFg S-M3 with data Augmentation



# Interpretability Results - Oxford IIIt Pet I





# Interpretability Results - Oxford IIIt Pet II

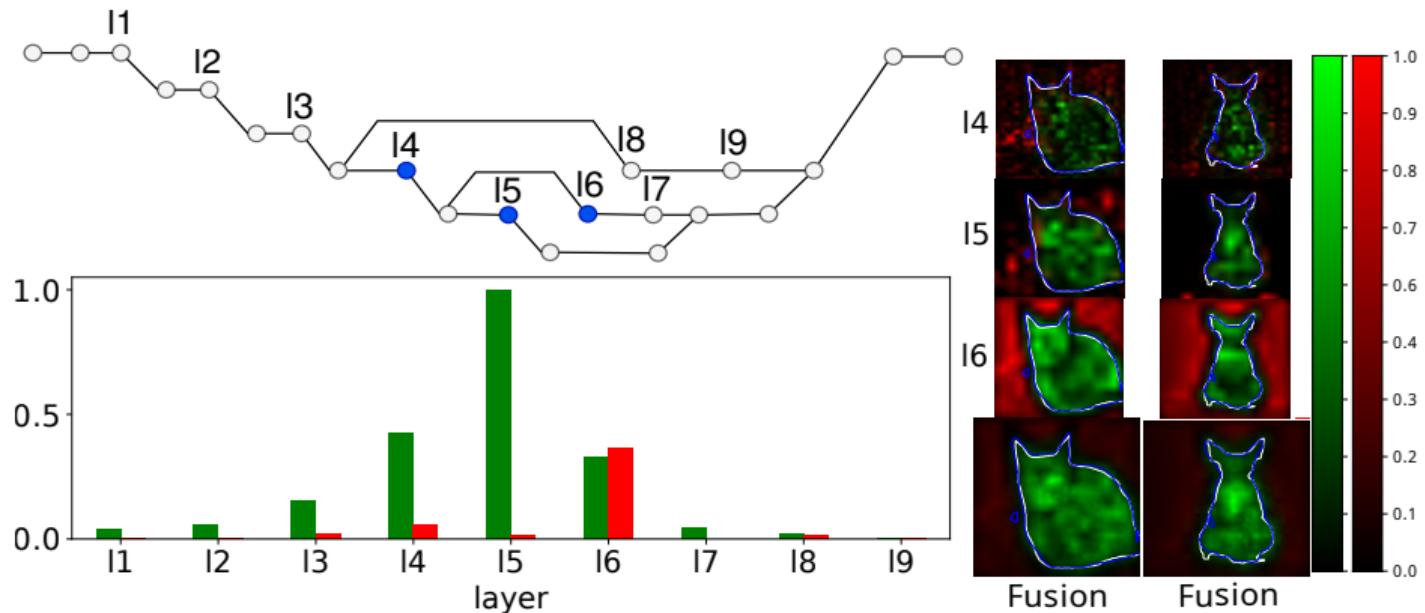


Figure: FCN S-M1



# Interpretability Results - Oxford IIIt Pet III

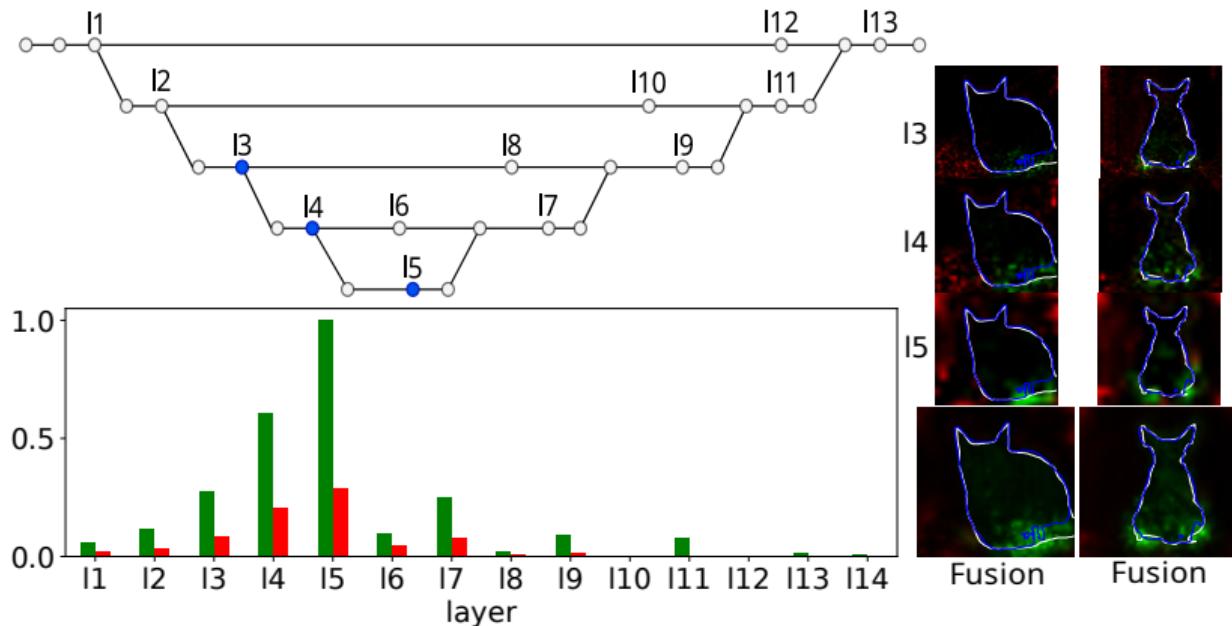


Figure: ResUNet S-M3



# Interpretability Results - Oxford IIIt Pet IV

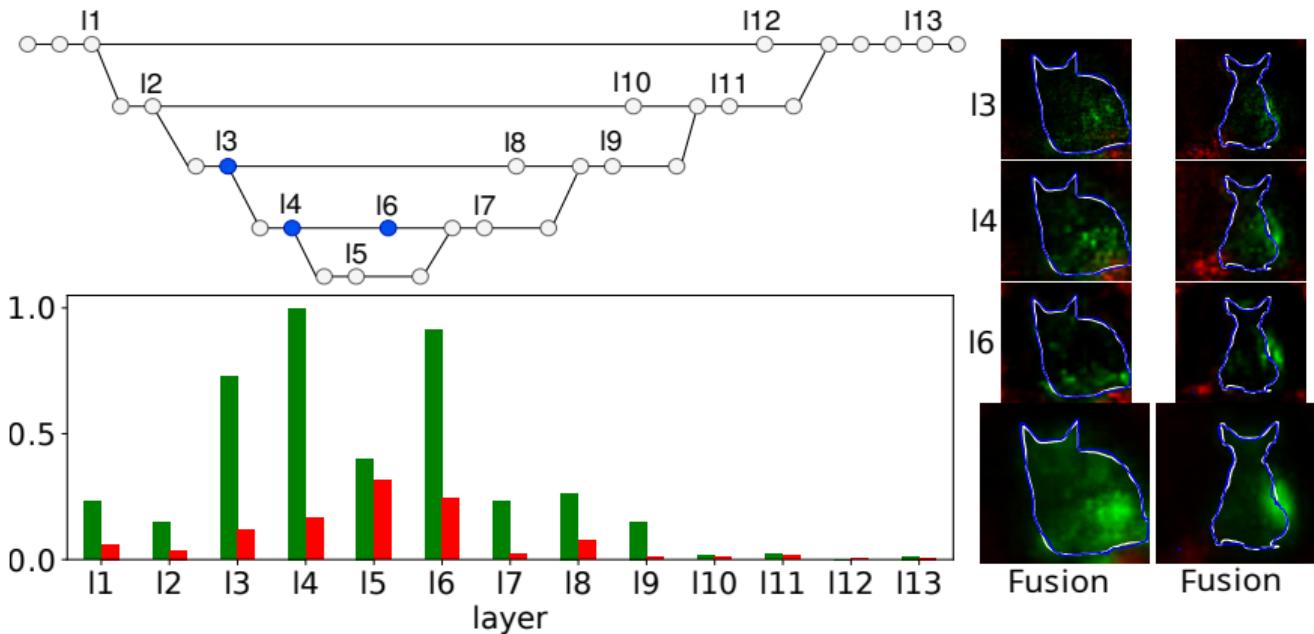


Figure: U-Net S-M1