Superpixel Correlation for Explainable Image Classification

The 3rd World Conference on eXplainable Artificial Intelligence, XAI 2025

Vahidin Hasic
Faculty of Electrical Engineering
University of Sarajevo, Sarajevo, Bosnia and Herzegovina

Amar Halilovic
Institute of Artificial Intelligence
Ulm University, Ulm, Germany

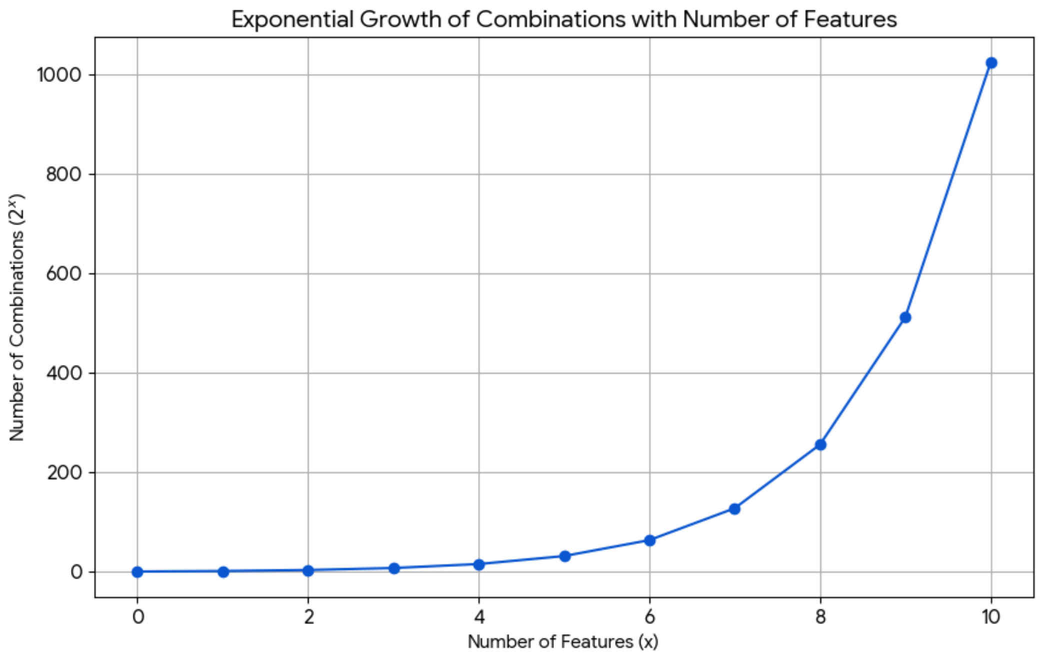
Senka Krivic
Faculty of Electrical Engineering
University of Sarajevo, Sarajevo, Bosnia and Herzegovina

Istanbul, Turkey

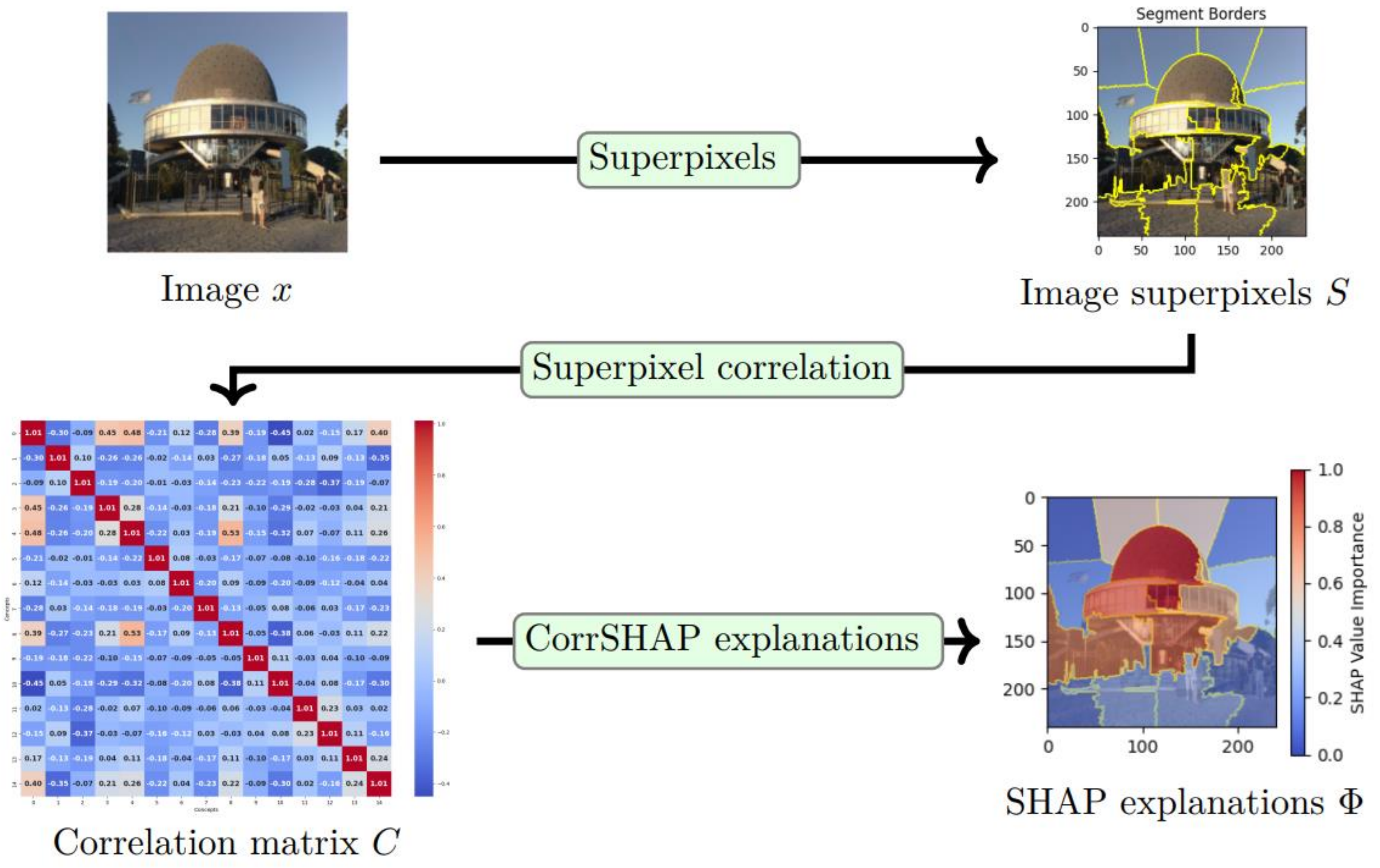
09-11 July, 2025

Introduction

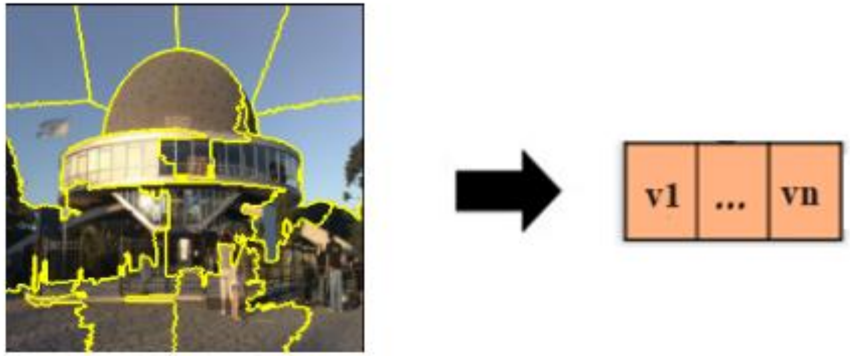
- Motivation:
  - SHAP:
    - Pixel-level explanations
    - Exponential computational complexity
  - Image superpixels are correlated
- Our contribution:
  - Novel SHAP approximation method CorrSHAP:
    - High-level superpixel explanations
    - Fast execution using novel superpixel correlation approach



Correlation SHAP (CorrSHAP)



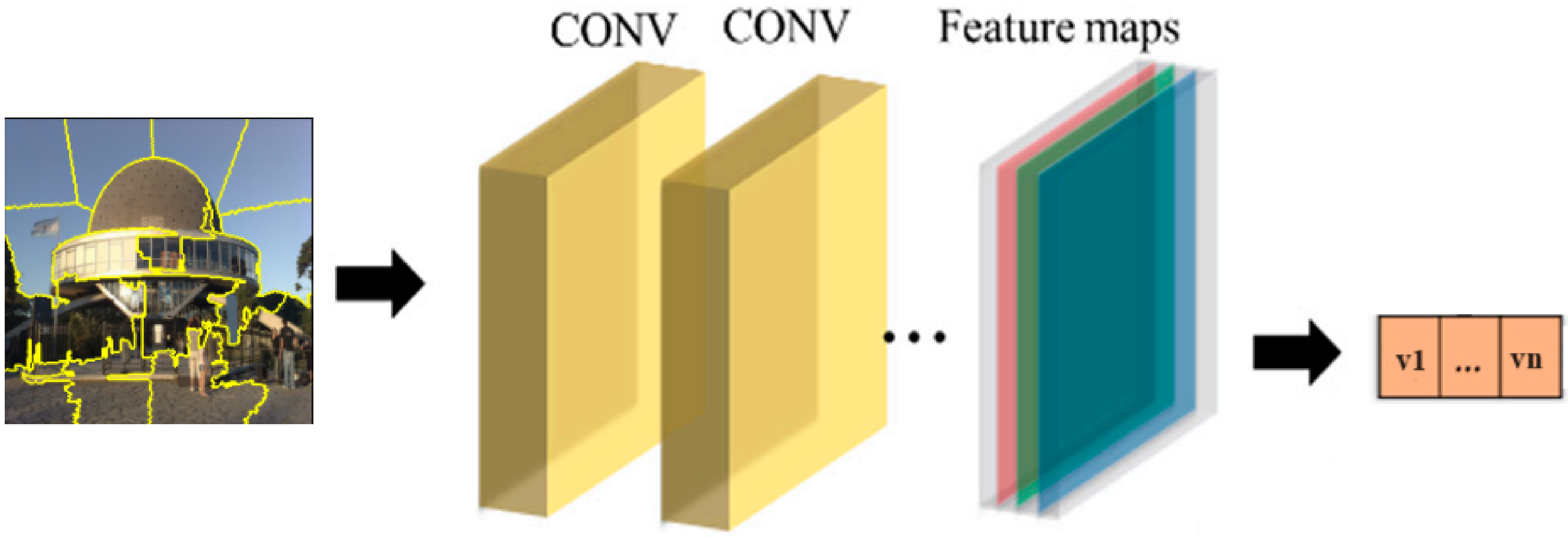
Raw Pixel Vectorization



$$v_i = \text{flatten}(s_i) = \text{flatten}(x \odot m_i)$$

$$v_i \in \mathbb{R}^d, x \in \mathbb{R}^{H \times W \times K}, m_i \in \{0,1\}^{H \times W \times K}, d = H \cdot W \cdot K.$$

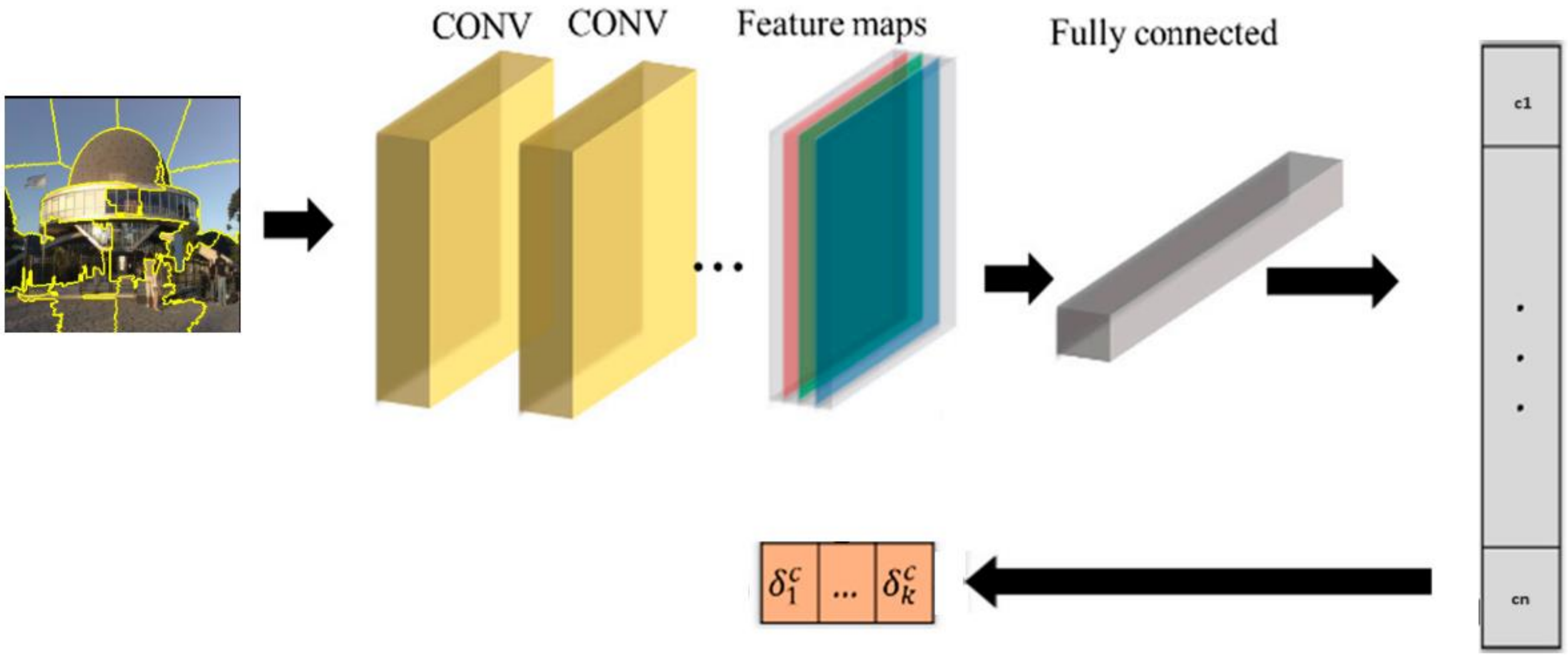
Feature Map Vectorization



$$v_i = \text{flatten}(g(s_i))$$

$$f(x) = h(g(x)), g : \mathcal{X} \to \mathcal{Z}, h : \mathcal{Z} \to \mathcal{Y}.$$

Gradient Vectorization



$$v_i = \nabla_\theta f_\theta(s_i).$$

CorrSHAP

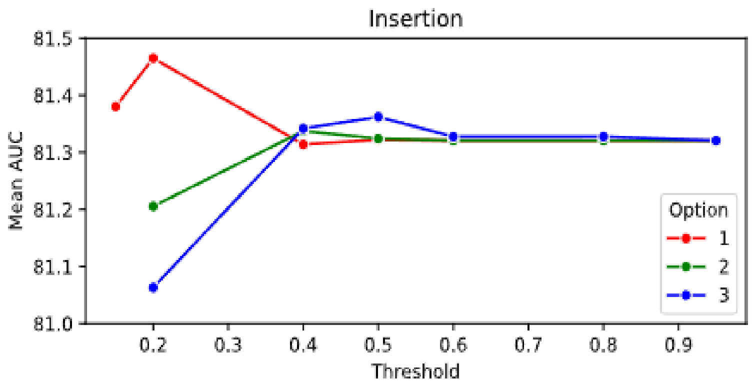
Algorithm 1 Correlation SHAP (CorrSHAP)

Input: Model f, Image x, ImageSegmentation q, Option ω
Output: SHAP superpixel attribution values $\hat{\phi}$

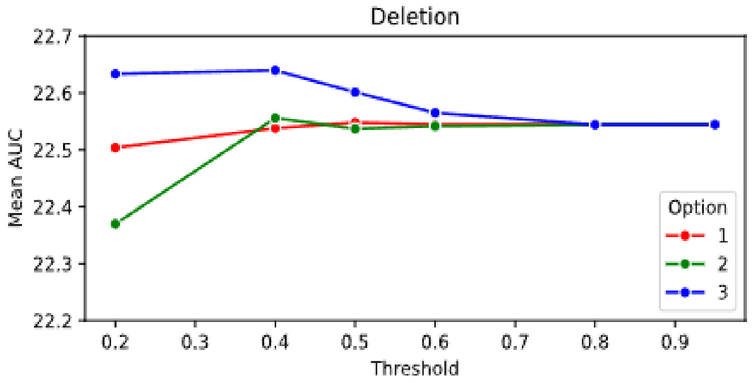
$g \leftarrow \text{FeatureExtractor}(f)$ $\{f(x) = h(g(x))\}$
$\mathcal{N} \leftarrow q(x)$ {Image segmentation}
if $\omega = 1$ **then**
    $V \leftarrow \text{flatten}(\mathcal{N})$ {Raw pixel vectorization}
else if $\omega = 2$ **then**
    $V \leftarrow \text{flatten}(g(\mathcal{N}))$ {Feature map vectorization}
else if $\omega = 3$ **then**
    $V \leftarrow \nabla_\theta f_\theta(\mathcal{N})$ {Gradient vectorization}
end if
$V' \leftarrow \text{Centralize(V)}$ {Vectors centralization}
$C = \{\cos(\mathbf{v}'_i, \mathbf{v}'_j) \mid \mathbf{v}'_i, \mathbf{v}'_j \in V'\}$ {Superpixel correlation matrix calculation}
$\hat{\phi} \leftarrow SHAP(\mathcal{N}, C)$ {Calculate SHAP values}
return $\hat{\phi}$

Ablation study

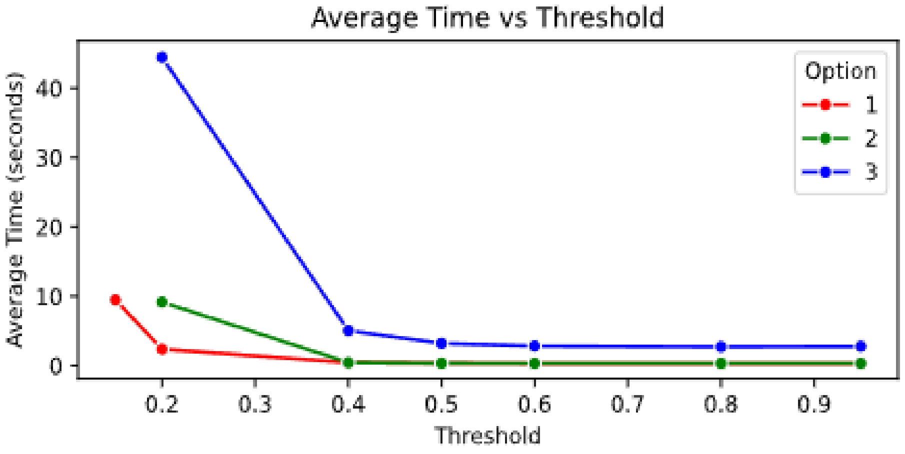
Option 1 – raw pixel vectorization
Option 2 – feature map vectorization
Option 3 – gradient vectorization



(a) Mean AUC insertion performance depending on threshold.

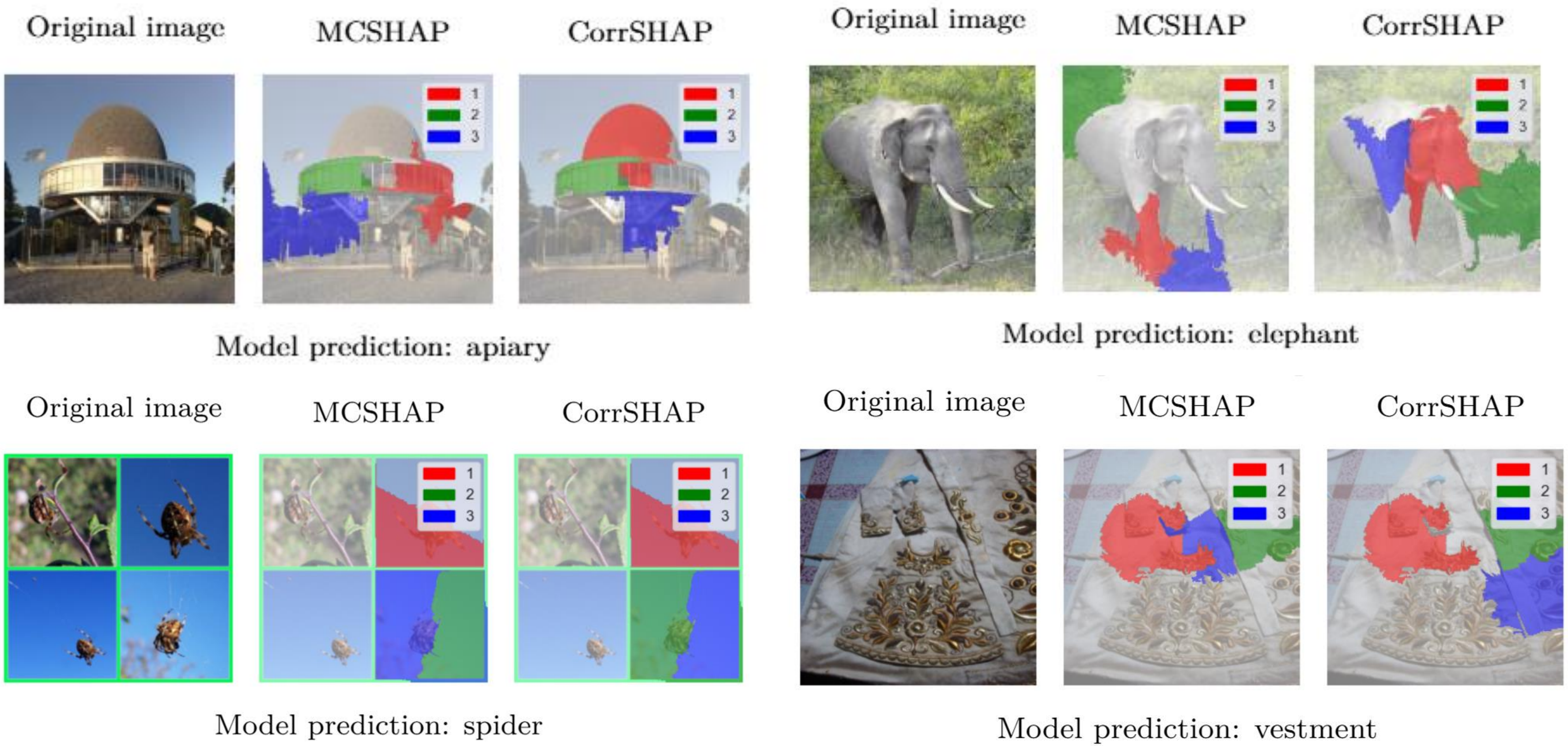


(b) Mean AUC deletion performance depending on threshold.



(c) Execution time depending on threshold.

Qualitative Results



Quantitative Results

- Speedup of 55x
- Higher faithfulness
- Vectorization approach 1 achieved the best performance overall

Area Under the Curve (AUC) Insertion ↑					
Model	Superpixels	CorrSHAP 1	CorrSHAP 2	CorrSHAP 3	MCSHAP
MobileNet-v2	Quickshift	80.4	80.37	80.29	**80.89**
	SLIC	**78.12**	78.13	78.15	77.79
ResNet-18	Quickshift	60.21	60.23	60.21	**61.60**
	SLIC	54.66	54.65	54.65	**55.41**
ResNet-50	Quickshift	**82.63**	82.61	82.63	82.27
	SLIC	**81.20**	81.20	81.20	80.66
ViT-b16	Quickshift	80.83	80.83	80.74	**81.84**
	SLIC	76.36	76.33	76.41	**76.82**
Area Under the Curve (AUC) Deletion ↓					
MobileNet-v2	Quickshift	**20.14**	20.14	20.16	20.93
	SLIC	**19.48**	19.48	19.48	21.40
ResNet-18	Quickshift	8.25	8.25	8.25	**8.03**
	SLIC	9.06	9.06	9.06	**9.01**
ResNet-50	Quickshift	**22.79**	22.79	22.79	24.16
	SLIC	**22.36**	22.36	22.39	23.79
ViT-b16	Quickshift	17.20	17.17	17.13	**16.86**
	SLIC	20.53	20.53	20.53	**20.48**
Execution Time (seconds) ↓					
MobileNet-v2	Quickshift	**0.42**	0.54	1.15	16.13
	SLIC	**0.74**	0.89	1.86	14.98
ResNet-18	Quickshift	**0.43**	0.50	1.09	7.85
	SLIC	0.66	**0.51**	1.91	13.82
ResNet-50	Quickshift	**0.46**	0.58	2.76	25.16
	SLIC	**0.78**	0.93	5.93	36.49
ViT-b16	Quickshift	**0.40**	0.52	5.41	7.26
	SLIC	**0.73**	1.01	10.53	18.00

Conclusion

- Novel SHAP approximation method CorrSHAP:
  - High-level, user understandable explanations – utilising superpixels
  - Fast SHAP calculation – utilising novel superpixel correlation approach

- Future work:
  - Alternative correlation measures
  - Integrating the superpixel correlation idea in other XAI methods

Thank you!

Vahidin Hasic

Contact data:

Amar Halilovic

Contact data:

Senka Krivic

Contact data: