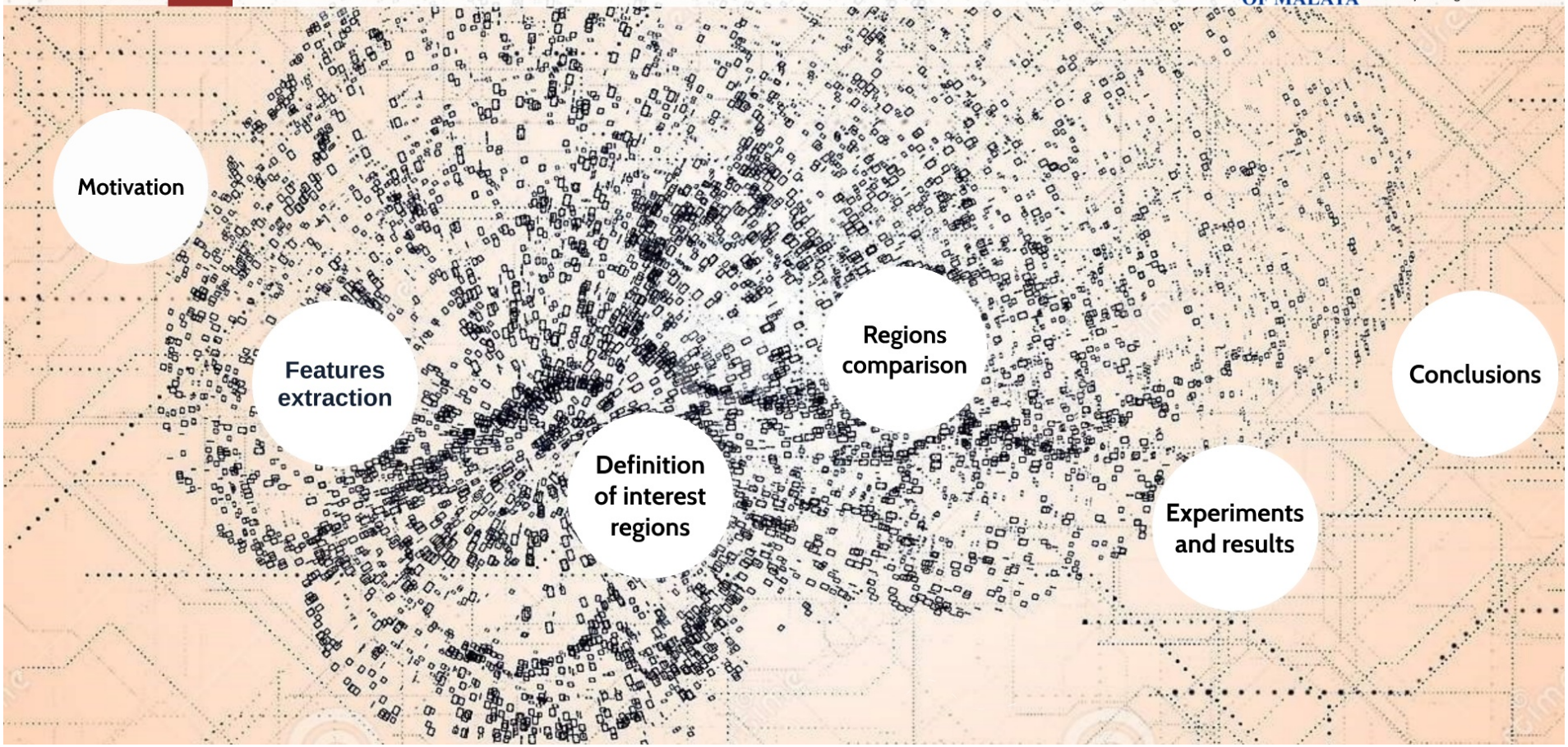
Marinella Iole Cadoni
Andrea Lagorio
Enrico Grosso
Computer Vision Laboratory
University of Sassari
Sassari, Italy
Email: maricadoni@uniss.it

From early biological models to CNNs: do they look where humans look?

Jia Huei, Tan
Chee Seng, Chan
Center of Image and Signal Processing
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Email: tanjiahuei@gmail.com



Our questions

Q1. Given an image, do CNNs look at the same regions humans look at?

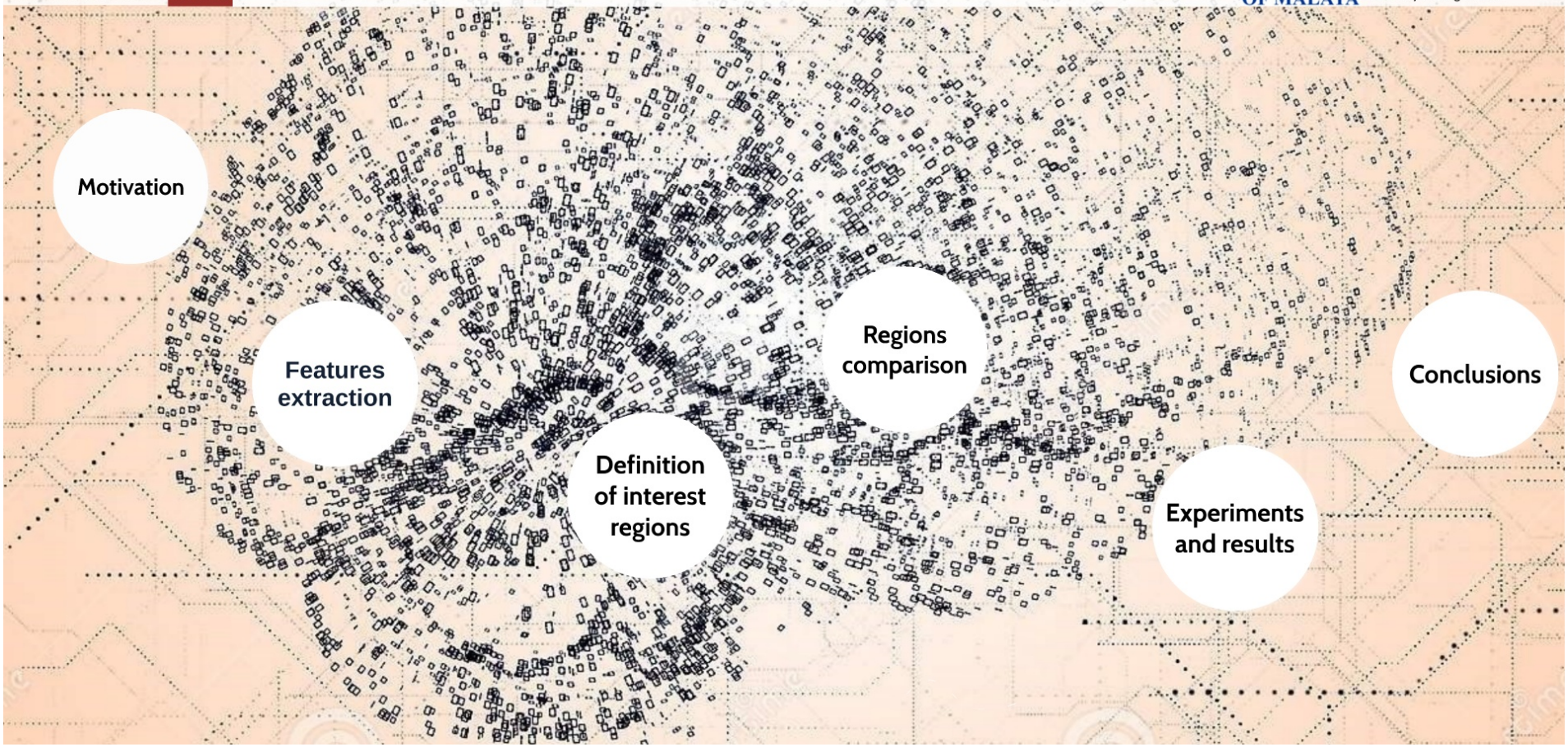
Q2. Is visual attention of recent networks more similar to humans attention than it was for earlier developed networks?

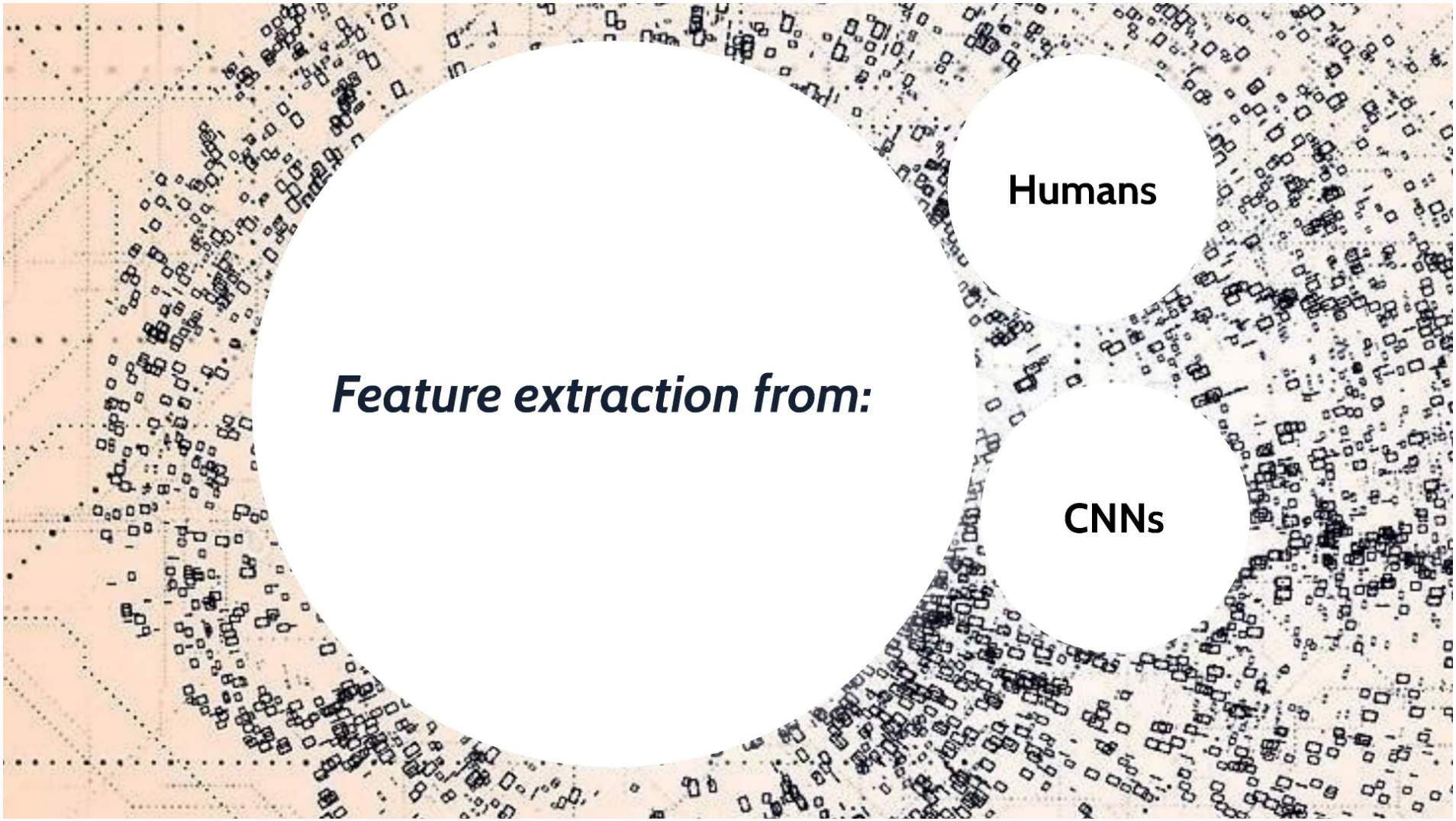
Q3. Is CNN's-human visual attention similarity related to CNN's classification performance?

Marinella Iole Cadoni
Andrea Lagorio
Enrico Grosso
Computer Vision Laboratory
University of Sassari
Sassari, Italy
Email: maricadoni@uniss.it

From early biological models to CNNs: do they look where humans look?

Jia Huei, Tan
Chee Seng, Chan
Center of Image and Signal Processing
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Email: tanjiahuei@gmail.com



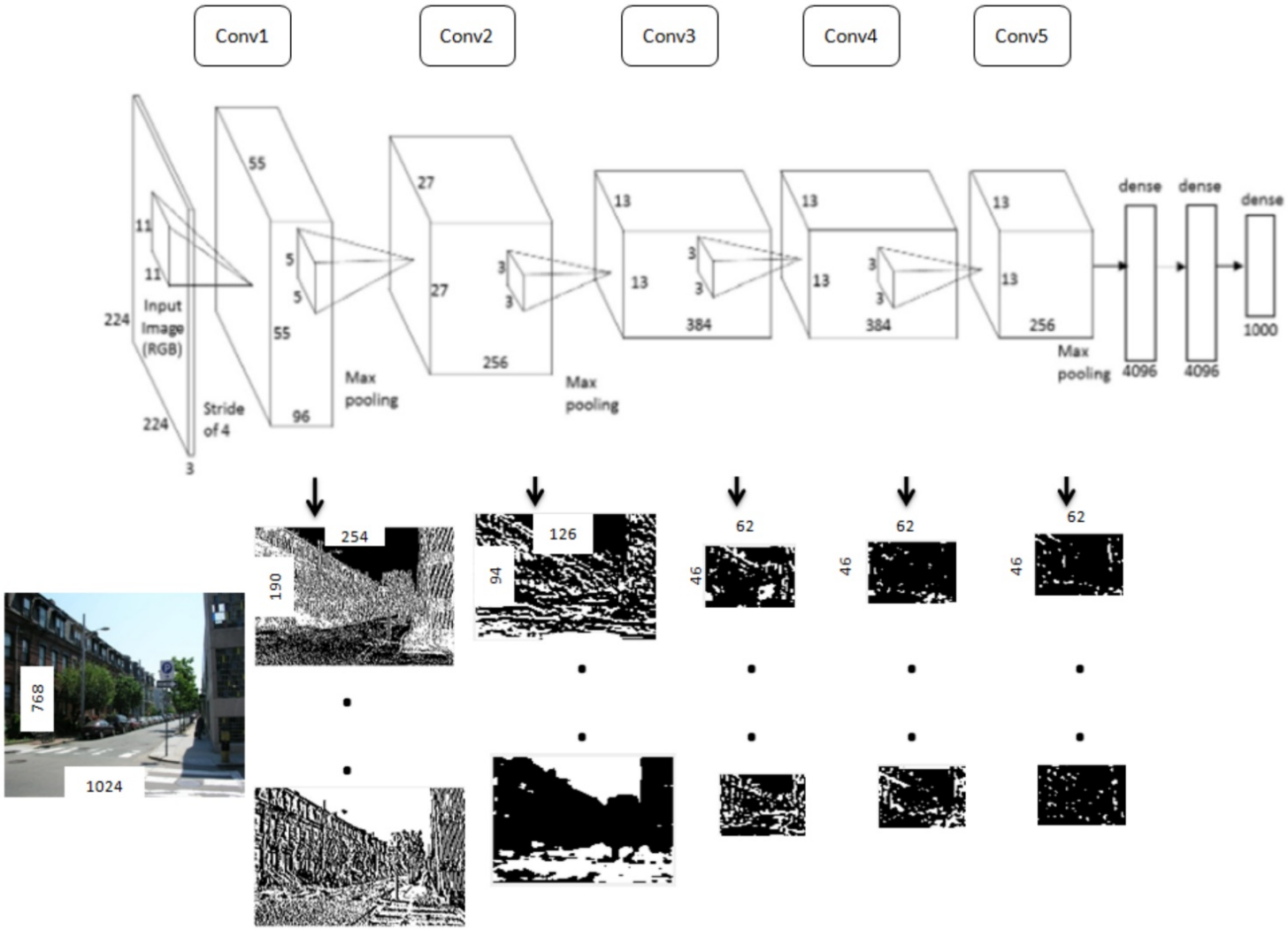

Feature extraction from:
Humans
CNNs

MIT Eye Tracking Database

- Gaze tracking paths (saccades and fixations)
- 15 users sitting in front of a screen
- 1003 images freely viewed
- 3 seconds per image with 1 second of gray screen in between

AlexNet

- 8 layers
- 5 convolutional
- 3 fully connected

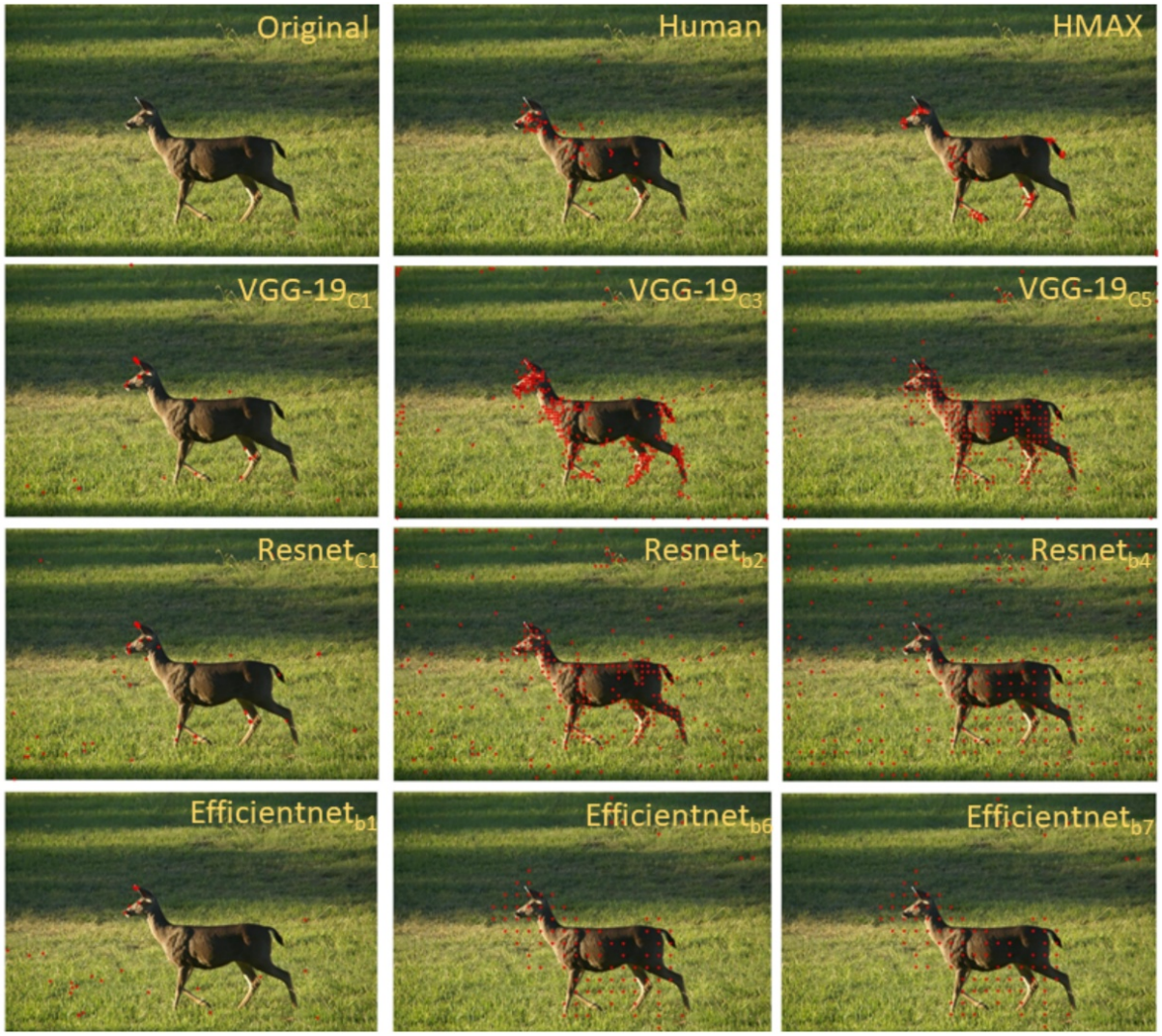


Example of human fixations and CNNs feature points.

Clockwise:

- original image
- humans fixations
- HMAX features

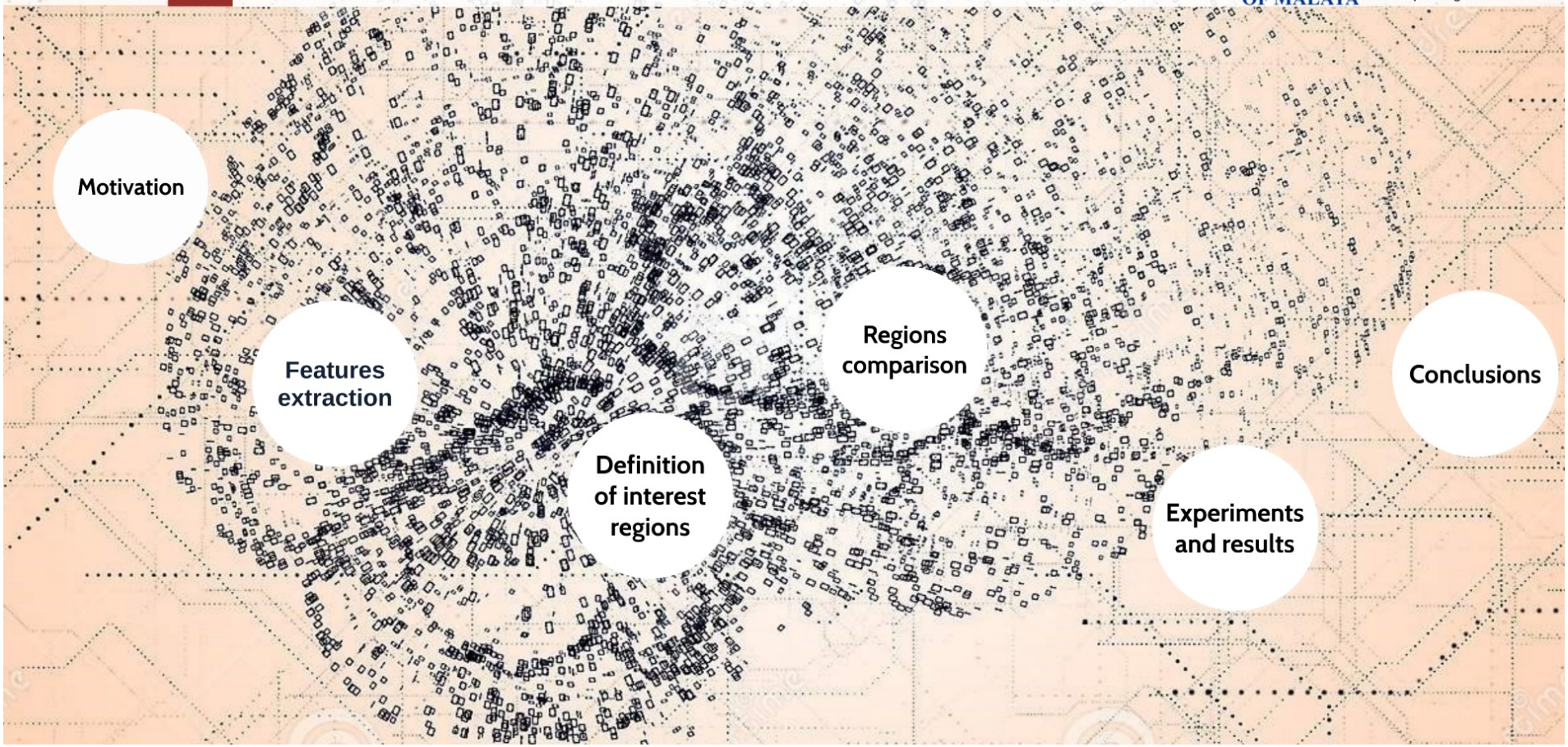
first, middle and last layers features of:
- VGG-19
- Resnet-50
- Efficientnet.



Marinella Iole Cadoni
Andrea Lagorio
Enrico Grosso
Computer Vision Laboratory
University of Sassari
Sassari, Italy
Email: maricadoni@uniss.it

From early biological models to CNNs: do they look where humans look?

Jia Huei, Tan
Chee Seng, Chan
Center of Image and Signal Processing
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Email: tanjiahuei@gmail.com

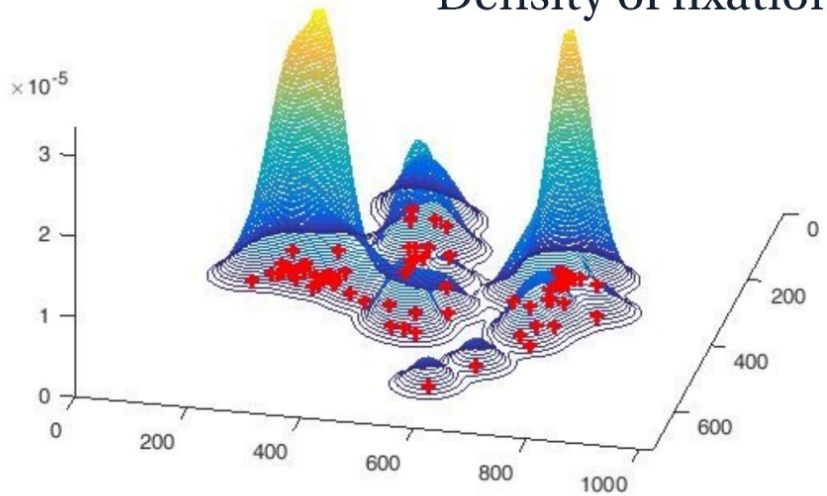


Regions of interest

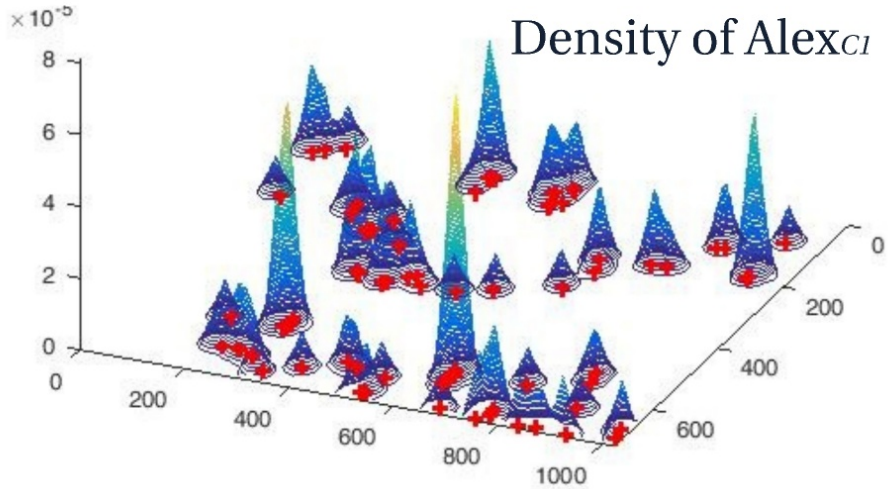
- Let I be an image and $F(I)$ a set of features.
- The region of interest is defined to be the support of the probability density function of the distribution of the feature points $F(I)$.
- The density $f_F(I)$ is estimated through a KDE (Kernel Density Estimation) based on a linear diffusion model proposed by Botev *et al.* [Kernel density estimation via diffusion, *The annals of statistics*, 2010]

Red points: Human fixations
Blue points: AlexNet$_{C1}$ features

Density of fixations



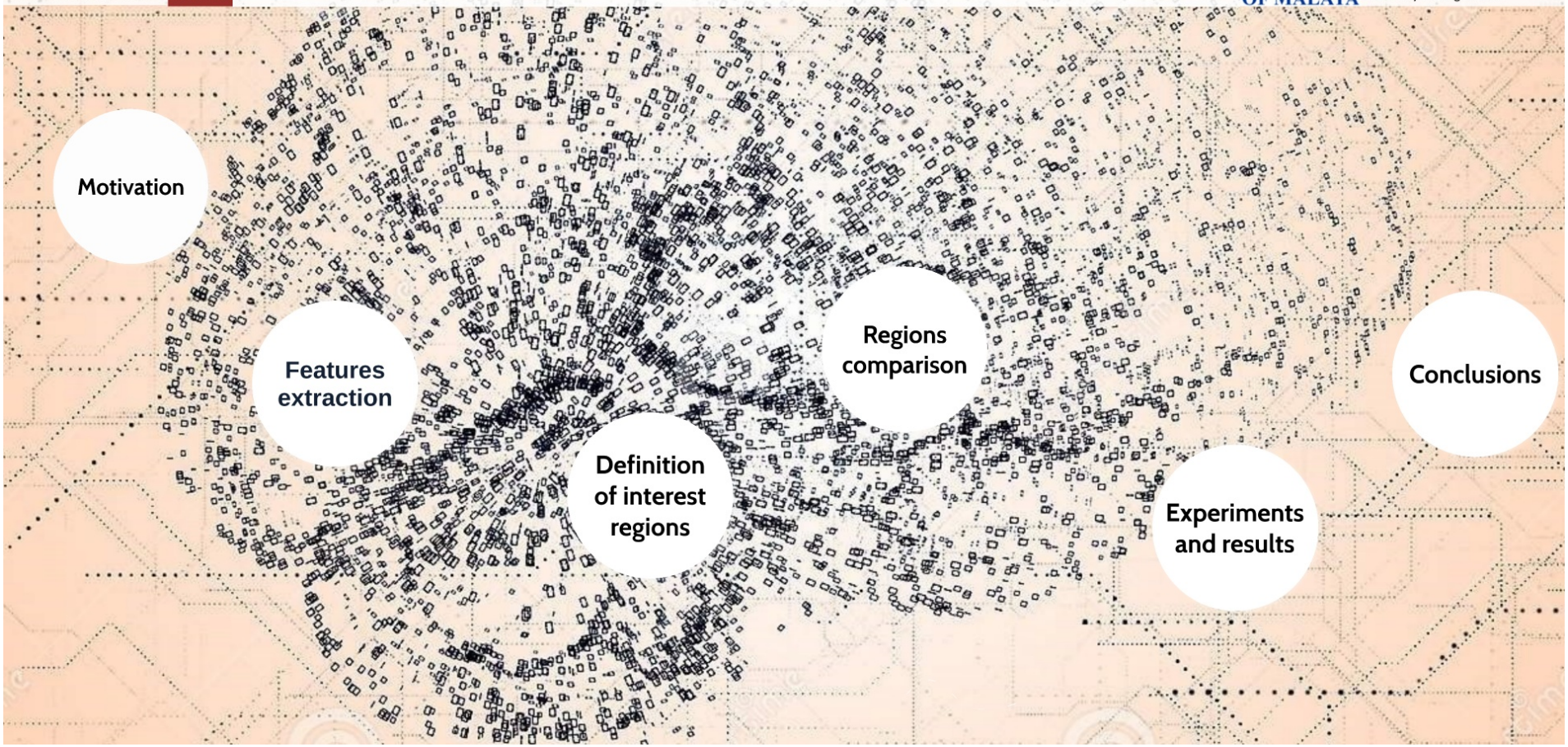
Density of Alex$_{C1}$

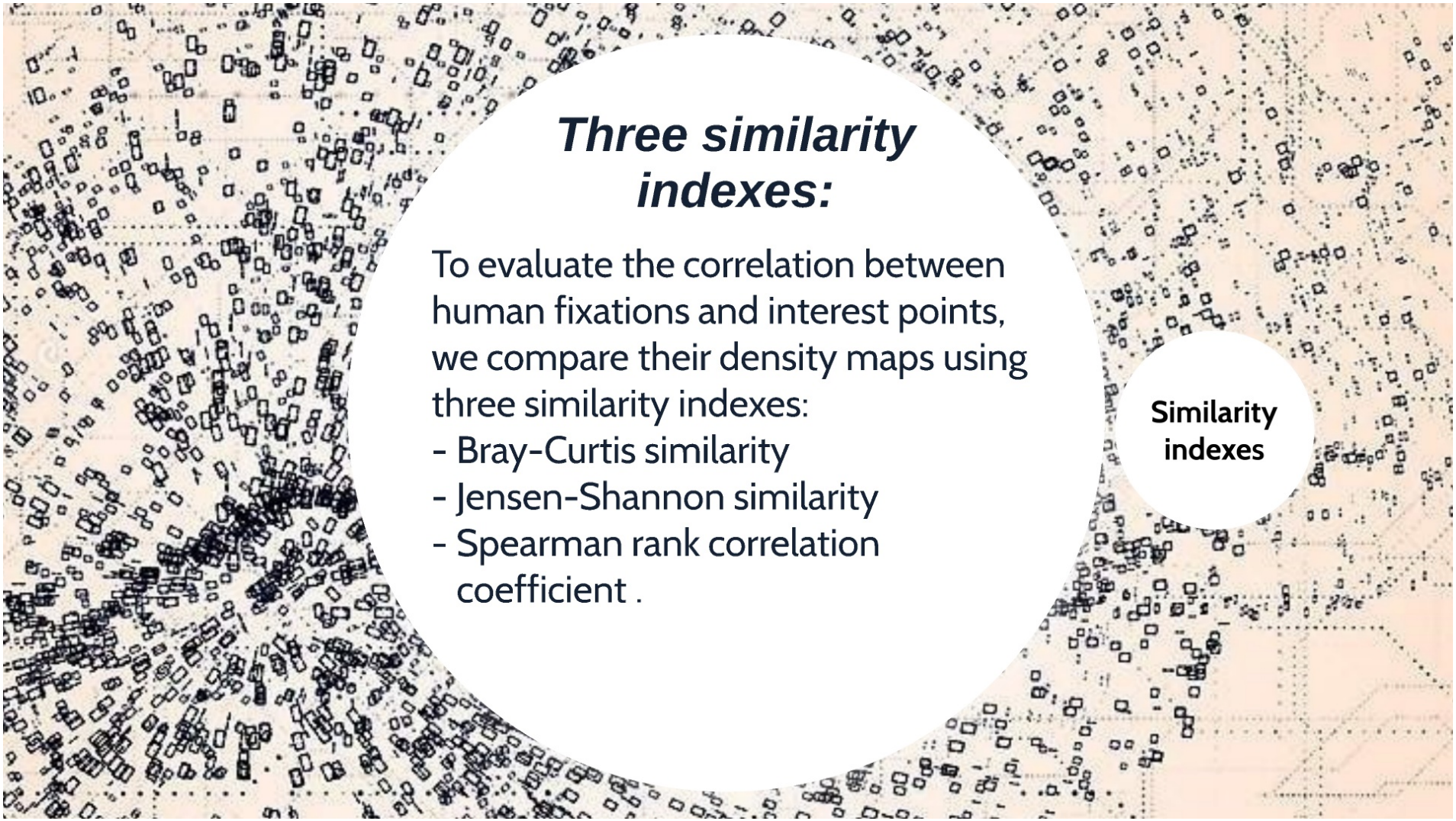


Marinella Iole Cadoni
Andrea Lagorio
Enrico Grosso
Computer Vision Laboratory
University of Sassari
Sassari, Italy
Email: maricadoni@uniss.it

From early biological models to CNNs: do they look where humans look?

Jia Huei, Tan
Chee Seng, Chan
Center of Image and Signal Processing
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Email: tanjiahuei@gmail.com





Three similarity indexes:

To evaluate the correlation between human fixations and interest points, we compare their density maps using three similarity indexes:

- Bray-Curtis similarity
- Jensen-Shannon similarity
- Spearman rank correlation coefficient .

Similarity indexes

Similarity indexes

Bray Curtis similarity: $$BC_{1,2} = 1 - \frac{\sum_{i=1}^{n} |f_1(\mathbf{x}_i) - f_2(\mathbf{x}_i)|}{\sum_{i=1}^{n} f_1(\mathbf{x}_i) + f_2(\mathbf{x}_i)}$$

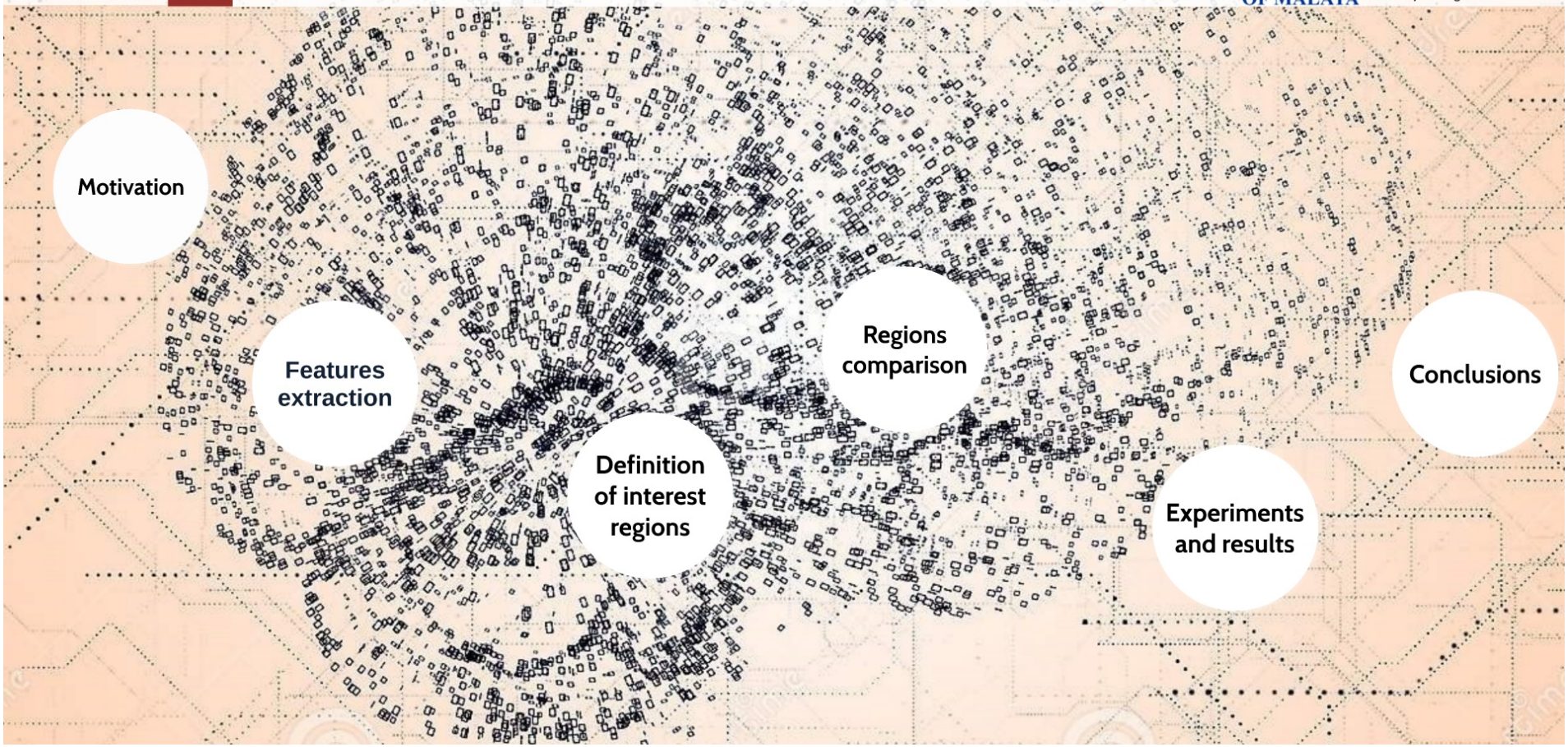
Jensen Shannon Divergence: $$JSD_{1,2} = \sum_{\mathbf{x}_i} (f_1(\mathbf{x}_i) - f_2(\mathbf{x}_i)) \log \frac{f_1(\mathbf{x}_i)}{f_2(\mathbf{x}_i)}$$

Jensen Shannon similarity: $$JS_{1,2} = 1 - JSD_{1,2}$$

Marinella Iole Cadoni
Andrea Lagorio
Enrico Grosso
Computer Vision Laboratory
University of Sassari
Sassari, Italy
Email: maricadoni@uniss.it

From early biological models to CNNs: do they look where humans look?

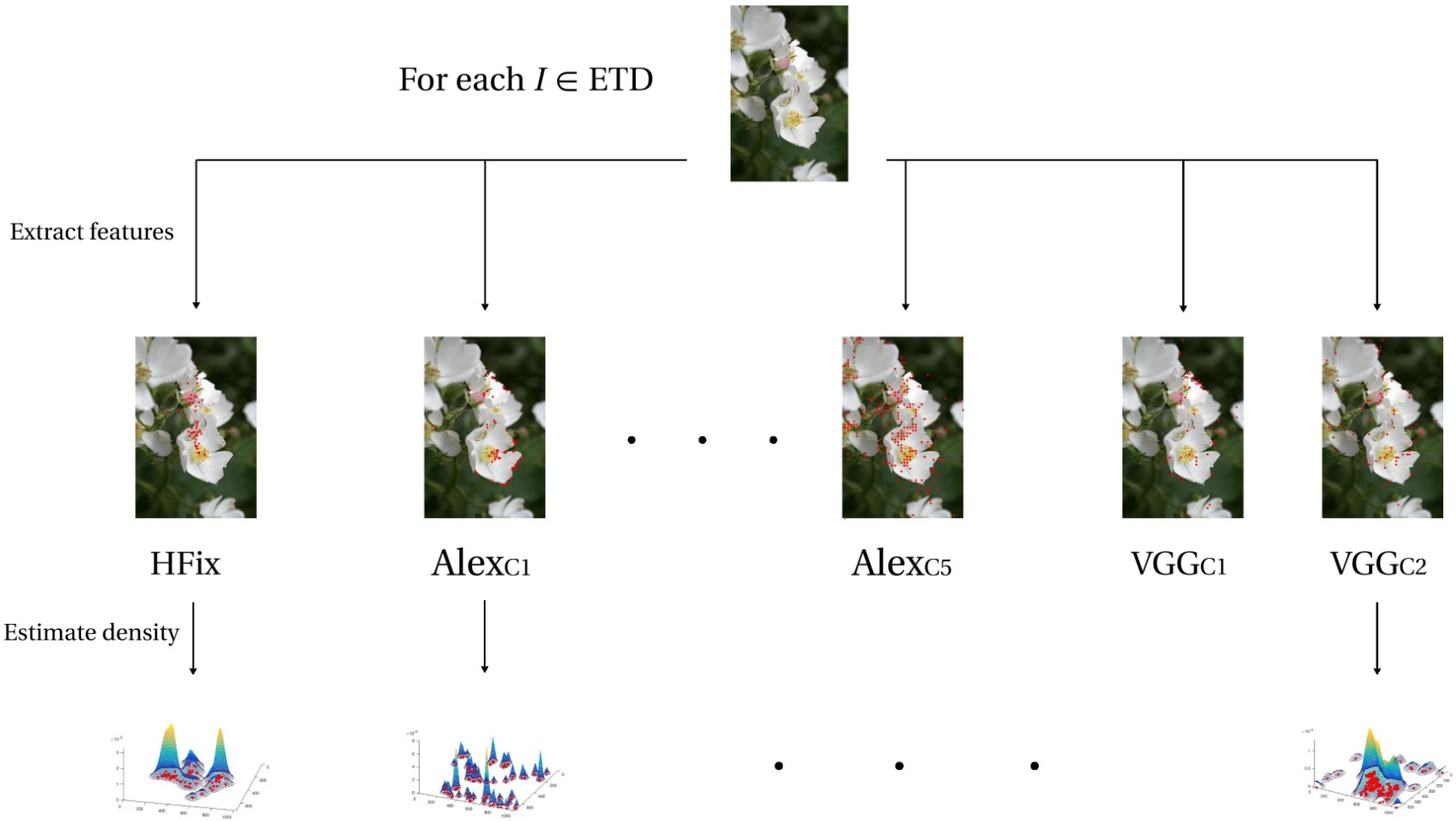
Jia Huei, Tan
Chee Seng, Chan
Center of Image and Signal Processing
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Email: tanjiahuei@gmail.com

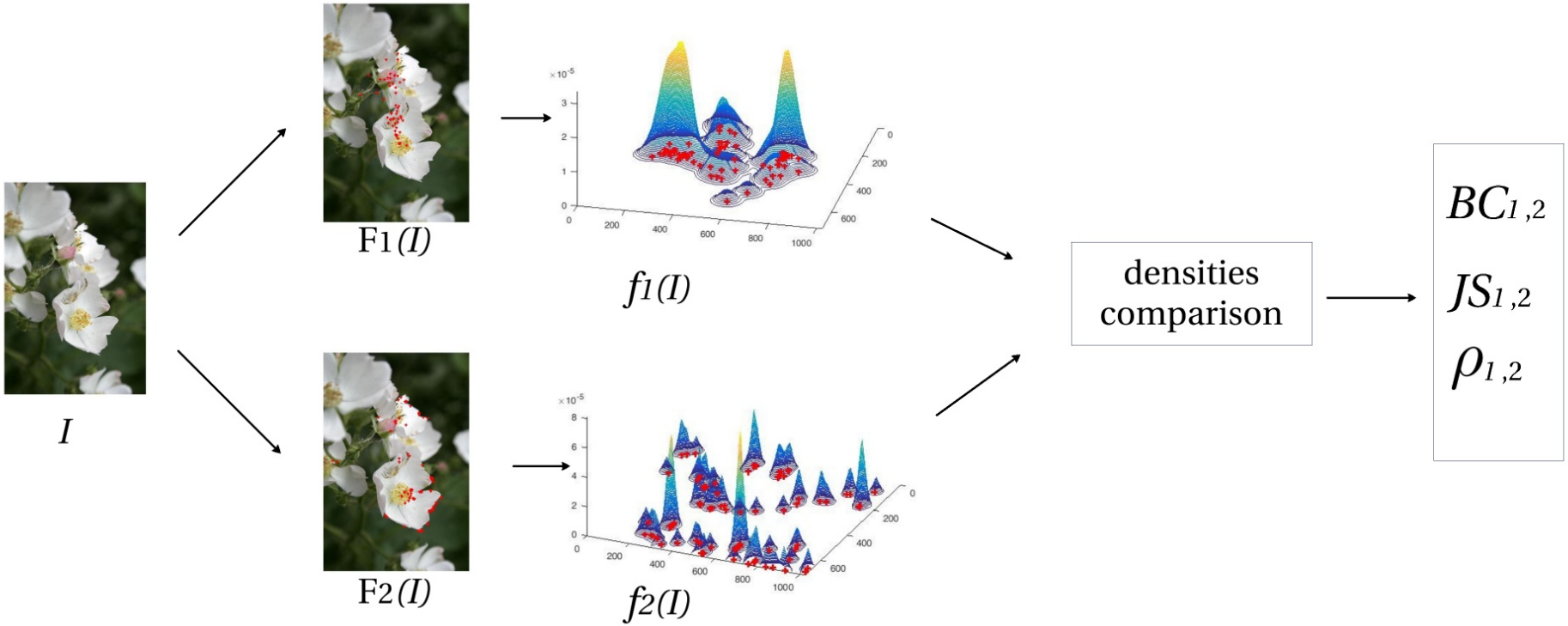


Experiments and results

Experimental protocol

Results


For each $I \in$ ETD
Extract features
HFix
AlexC1
AlexC5
VGGC1
VGGC2
Estimate density

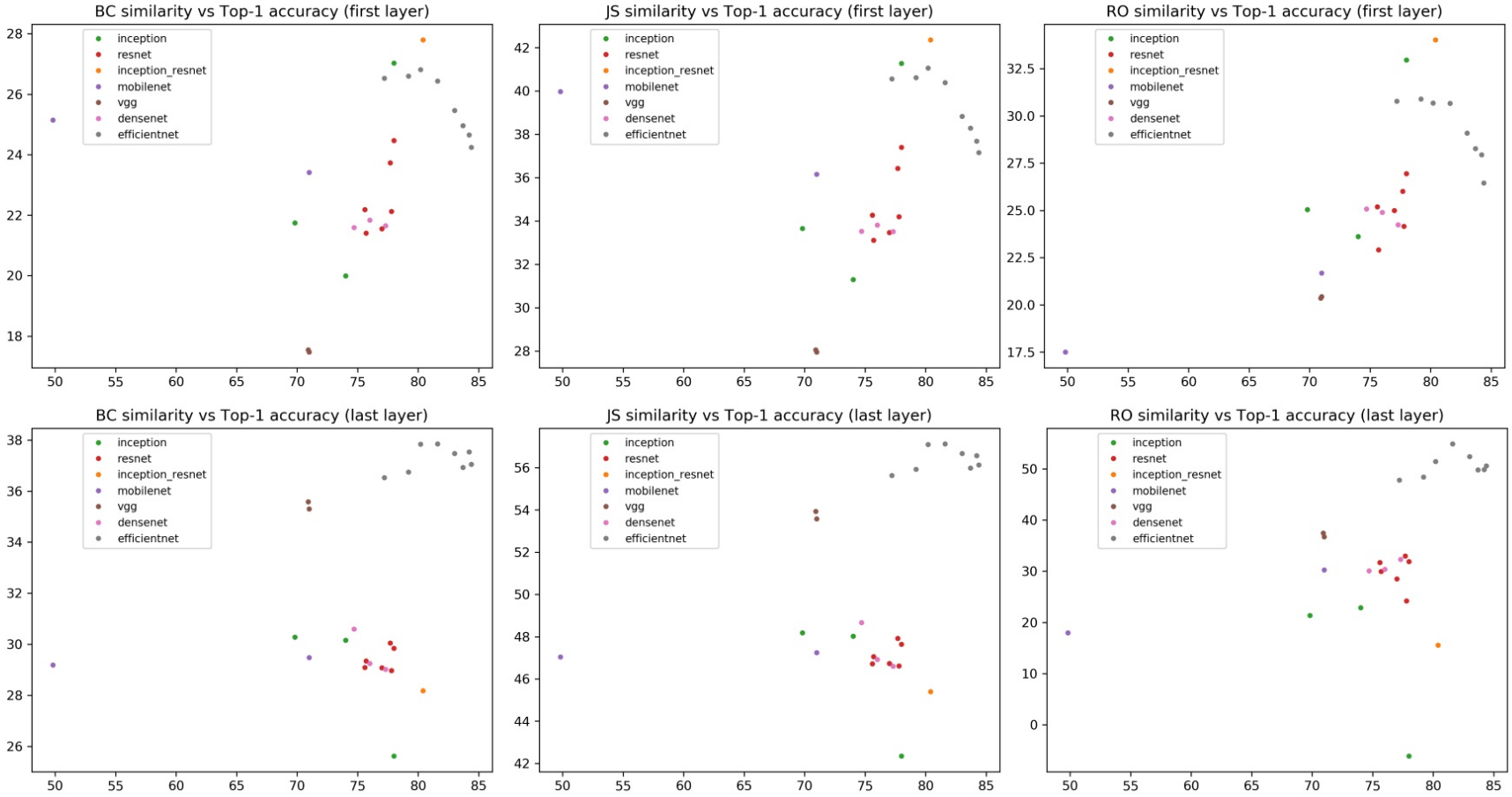

I
F1(I)
F2(I)
f1(I)
f2(I)
densities comparison
BC1,2
JS1,2
ρ1,2

Summary table of comparisons between human fixations and CNNs

Model	Layer	$BC_{f_{F_j} f_H}$	$JS_{f_{F_j} f_H}$	$\rho_{f_{F_j} f_H}$
		R=21.26%	R=34.11%	
HMAX		15.24%	24.81%	18.47
AlexNet	$C1$	25.51%	38.95%	31.36
	$C2$	31.61%	48.59%	29.66
	$C3$	34.45%	53.05%	37.19
	$C4$	34.47%	52.94%	36.38
	$C5$	34.09%	52.49%	35.39
VGG-19	$C1$	14.03%	22.99%	14.47
	$C2$	20.51%	32.12%	21.94
	$C3$	20.24%	47.91%	20.34
	$C4$	33.46%	50.05%	33.91
	$C5$	35.18%	53.44%	37.16
ResnetV2-50	$C1$	22.18%	**34.27%**	25.20
	$b1$	32.51%	49.42%	30.24
	$b2$	35.33%	54.24%	41.12
	$b3$	29.42%	47.09%	36.81
	$b4$	29.09%	46.71%	31.75

Model	Layer	$BC_{f_{F_j} f_H}$	$JS_{f_{F_j} f_H}$	$\rho_{f_{F_j} f_H}$
		R=21.26%	R=34.11%	
InceptionV3	$c1$	27.03%	41.27%	32.96
	$c2$	26.51%	40.40%	28.68
	$c3$	**21.87%**	**33.38%**	21.34
	$c4$	**21.87%**	**33.53%**	19.49
	$c5$	**21.88%**	**34.93%**	12.99
	$c6$	27.54%	44.63%	5.53
	$c7$	25.62%	42.35%	-6.14
Densenet-201	$C1$	**21.66%**	**33.53%**	24.25
	$C2$	29.82%	45.24%	27.28
	$C3$	34.65%	52.82%	32.24
	$C4$	32.36%	50.92%	30.16
	$C5$	29.02%	46.60%	32.20
EfficientNet-b7	$b1$	28.27%	42.95%	36.69
	$b2$	34.85%	52.64%	40.12
	$b3$	32.42%	50.17%	32.59
	$b4$	31.21%	49.01%	28.63
	$b5$	33.37%	51.42%	31.15
	$b6$	37.71%	56.98%	52.33
	$b7$	37.37%	56.36%	44.92

Correlation between human similarity and classification performance. x-axis: performance, y-axis: human similarity.



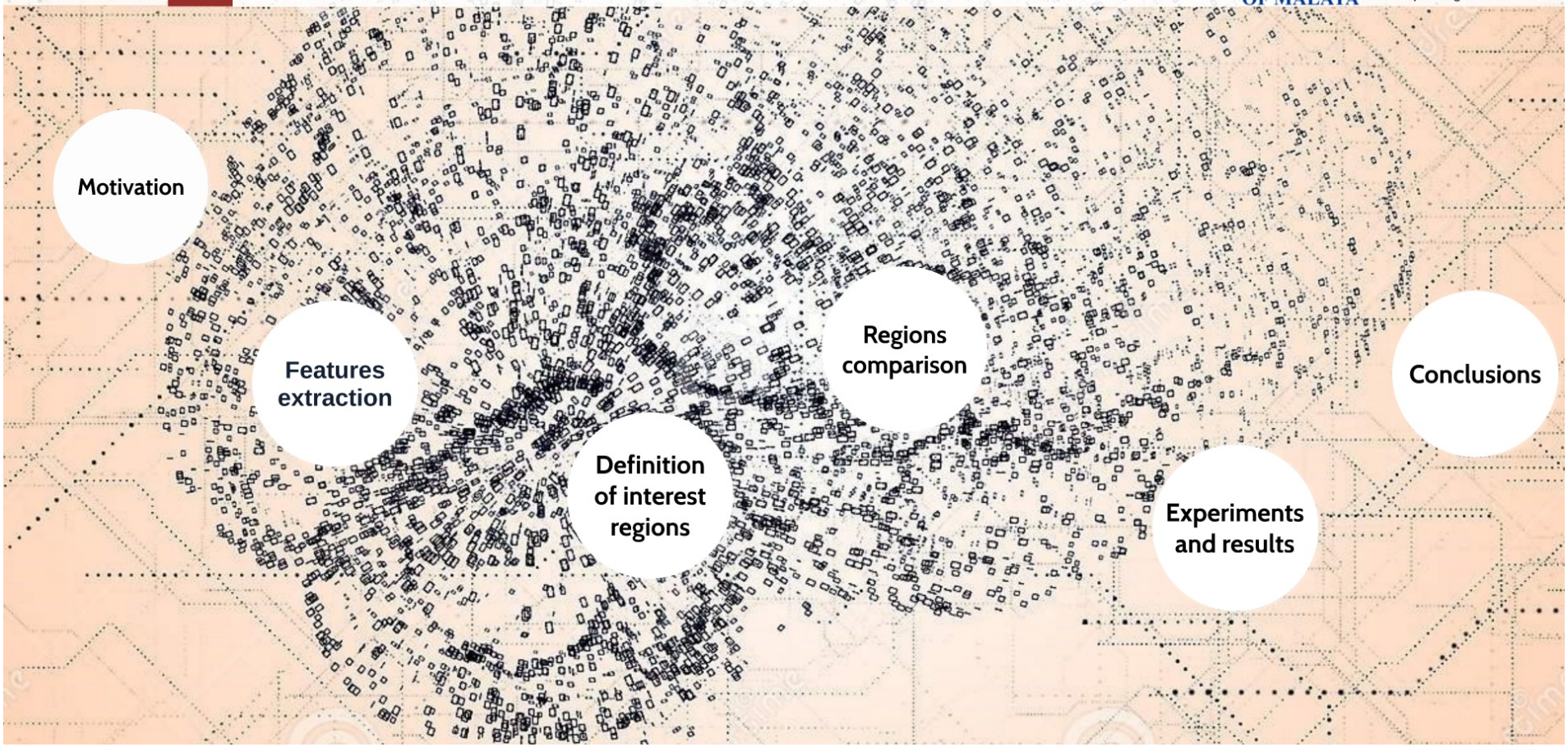
Spearman rank correlation test between human similarity and performance:

	Bray-Curtis	Jensen-Shannon	Spearman ρ
First layer	0.63	0.61	0.76
Top layer	0.38	0.37	0.54

Marinella Iole Cadoni
Andrea Lagorio
Enrico Grosso
Computer Vision Laboratory
University of Sassari
Sassari, Italy
Email: maricadoni@uniss.it

From early biological models to CNNs: do they look where humans look?

Jia Huei, Tan
Chee Seng, Chan
Center of Image and Signal Processing
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Email: tanjiahuei@gmail.com



Conclusions

Q1. *Do CNNs look at the same regions humans look at?*

A. **HMAX features and filters responses from shallow CNN's layers are not similar to human fixations, while filter responses of deeper layers are.**

Q2. *Is visual attention of recent networks more similar to humans attention than the attention of earlier developed networks?*

A. **It is if we consider early models such as HMAX, while the attention of networks such as AlexNet and VGG is more or less as similar to humans as that of more recent networks such as Densenet and Efficientnet.**

Q3. *Is CNNs-human visual attention similarity related to CNNs classification performance?*

A. **Yes it is: there is positive correlation between similarity and performance.**


danke
謝謝
ngiyabonga
teşekkür ederim
спасибо
dank je
thank you
gracias
tapadh leat
mauruuru
hvala
mochchakkeram
bedankt
dziękuję
go raibh maith agat
sagolun
sukriya
kop khun krap
obrigado
arigatō
takk
dakujem
grazie
мерси
terima kasih
감사합니다
merci
ευχαριστώ

Marinella Iole Cadoni
Andrea Lagorio
Enrico Grosso
Computer Vision Laboratory
University of Sassari
Sassari, Italy
Email: maricadoni@uniss.it

From early biological models to CNNs: do they look where humans look?

Jia Huei, Tan
Chee Seng, Chan
Center of Image and Signal Processing
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Email: tanjiahuei@gmail.com