Font Shape-to-Impression Translation

Masaya Ueda[*1]

Akisato Kimura[*2]

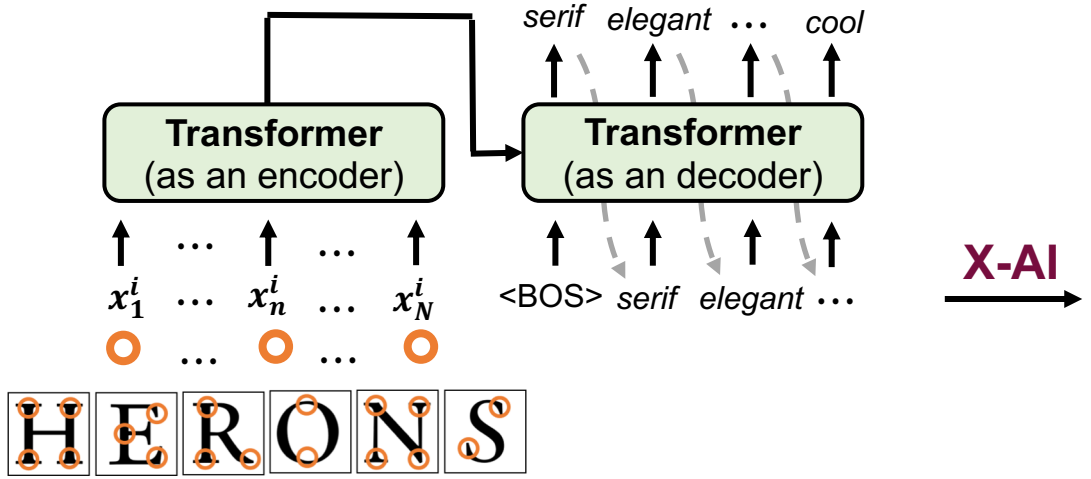
Seiichi Uchida[*1]

*1 Kyushu University, Fukuoka, Japan

*2 NTT Communication Science Laboratories, Kanagawa, Japan

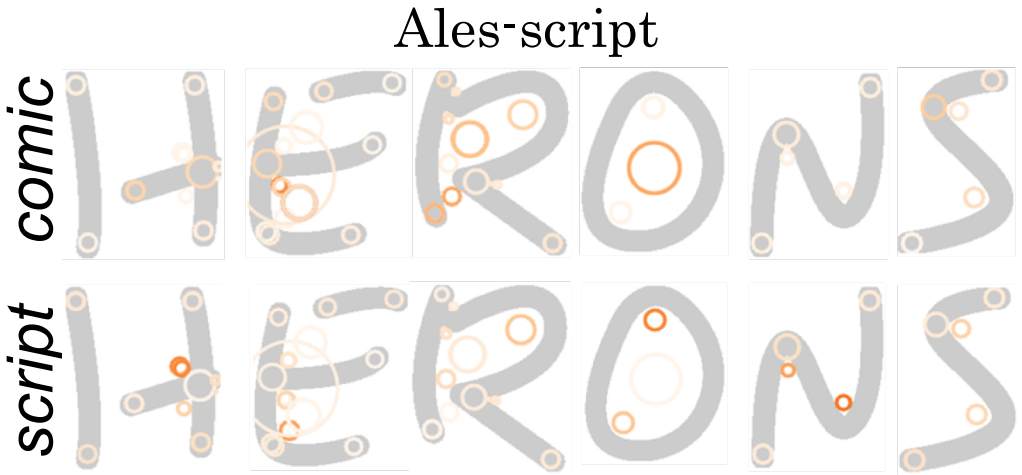
Key Takeaways

- **Contributions : New part-based font-impression analysis**
  - **Understand the important parts by Transformer and explainable-AI(X-AI) techniques**



Font Shape-to-Impression Translation

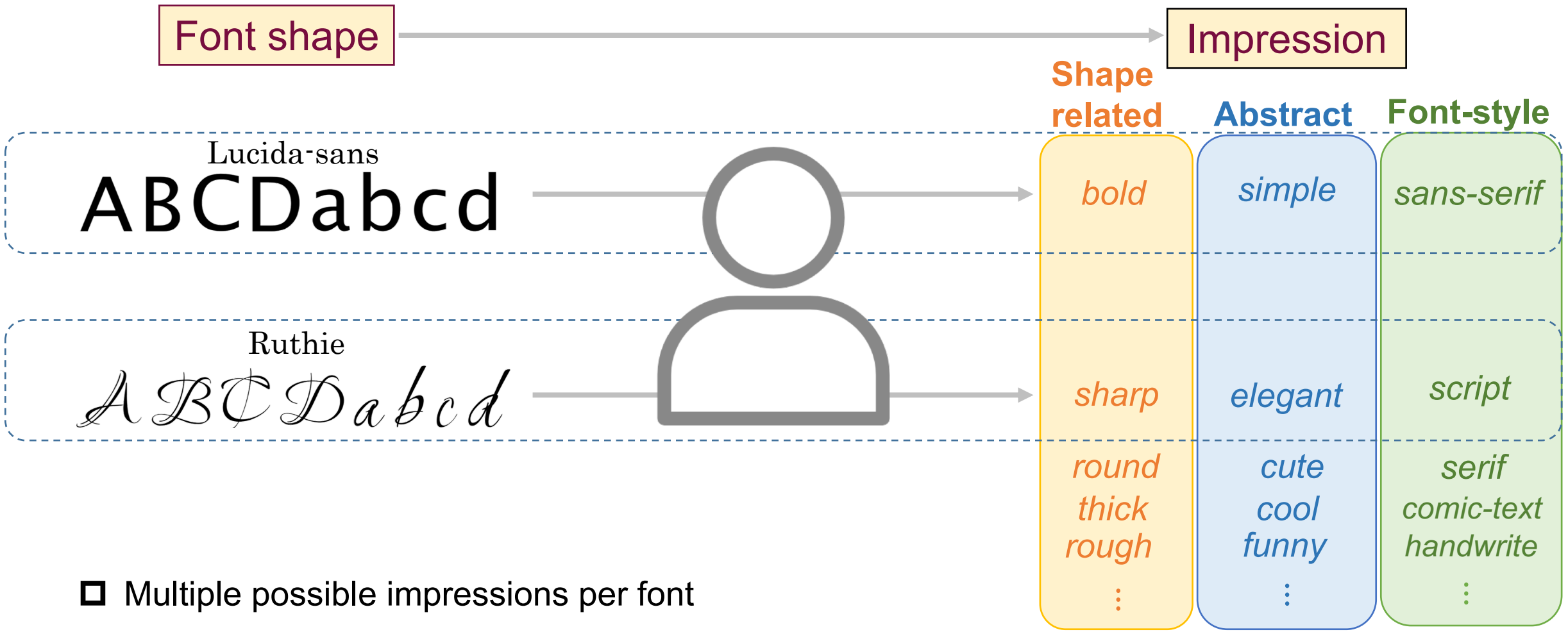
X-AI



Analysis of Important parts for impressions

- **Applications : Font recommendation and generation systems for specific impression**

Motivation : Different fonts have different impressions



- Multiple possible impressions per font
- Relationship between font shape and impression is unrevealed

Purpose : Analyze the relationship between font shape and impression

Font **local shape** =
- Stroke end
- Corner
- Curve

⇔ Impression

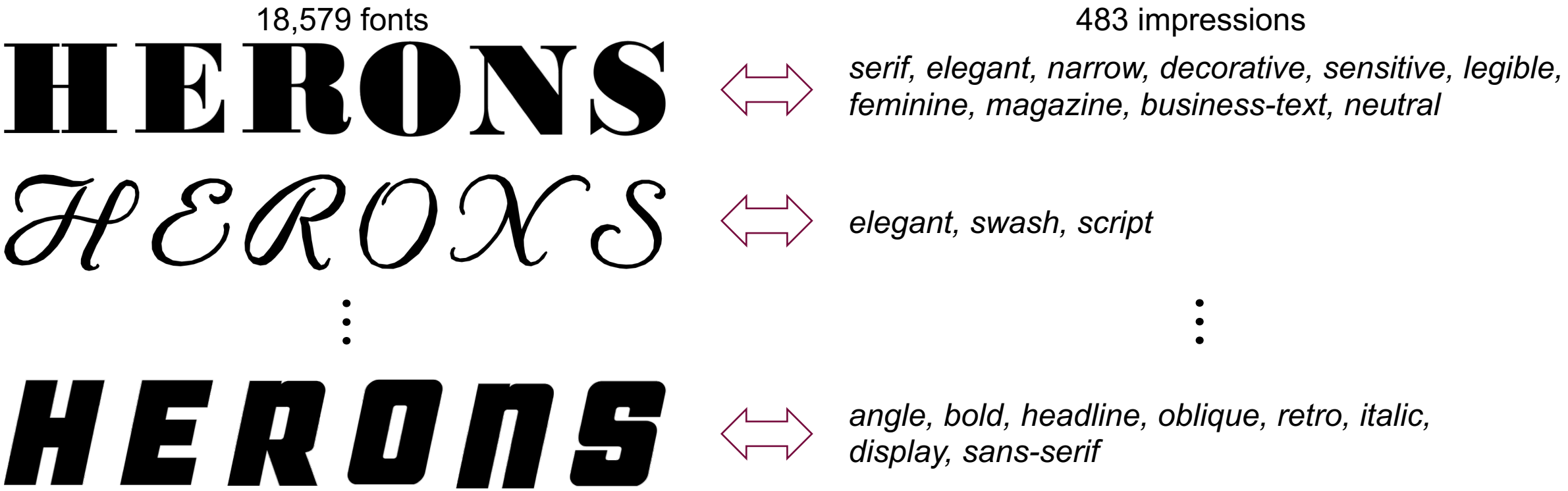
Century-old-style-sh

⇔
- *serif*
- *cool*
- *elegant*

- Whole shapes → Affects character classes such as "A" and "Z"
- Local shapes → Not affect character class and the decoration appears

Font-Impression Dataset [*]

- **18,815 fonts** and **1,824 impressions** dataset collected from MyFonts.com
- Use the impressions contained **in more than 100 fonts** and the fonts labeled with them



[*] Chen T. et al., ”Large-scale Tag-based Font Retrieval with Generative Feature Learning”, ICCV(2019)

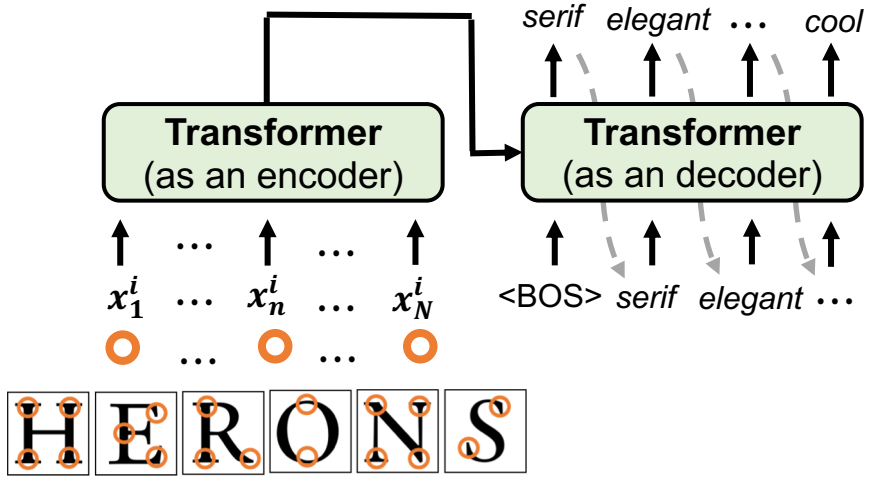
Methodology:

Shape-to-impression translation

Strategy : Local Shape × Transformer [*] × X-AI

① Local shape → Feature (SIFT[**]) vector (extracted from “HERONS” per a font)

② Output an arbitrary number of impression from a set of SIFT vectors by Transformer

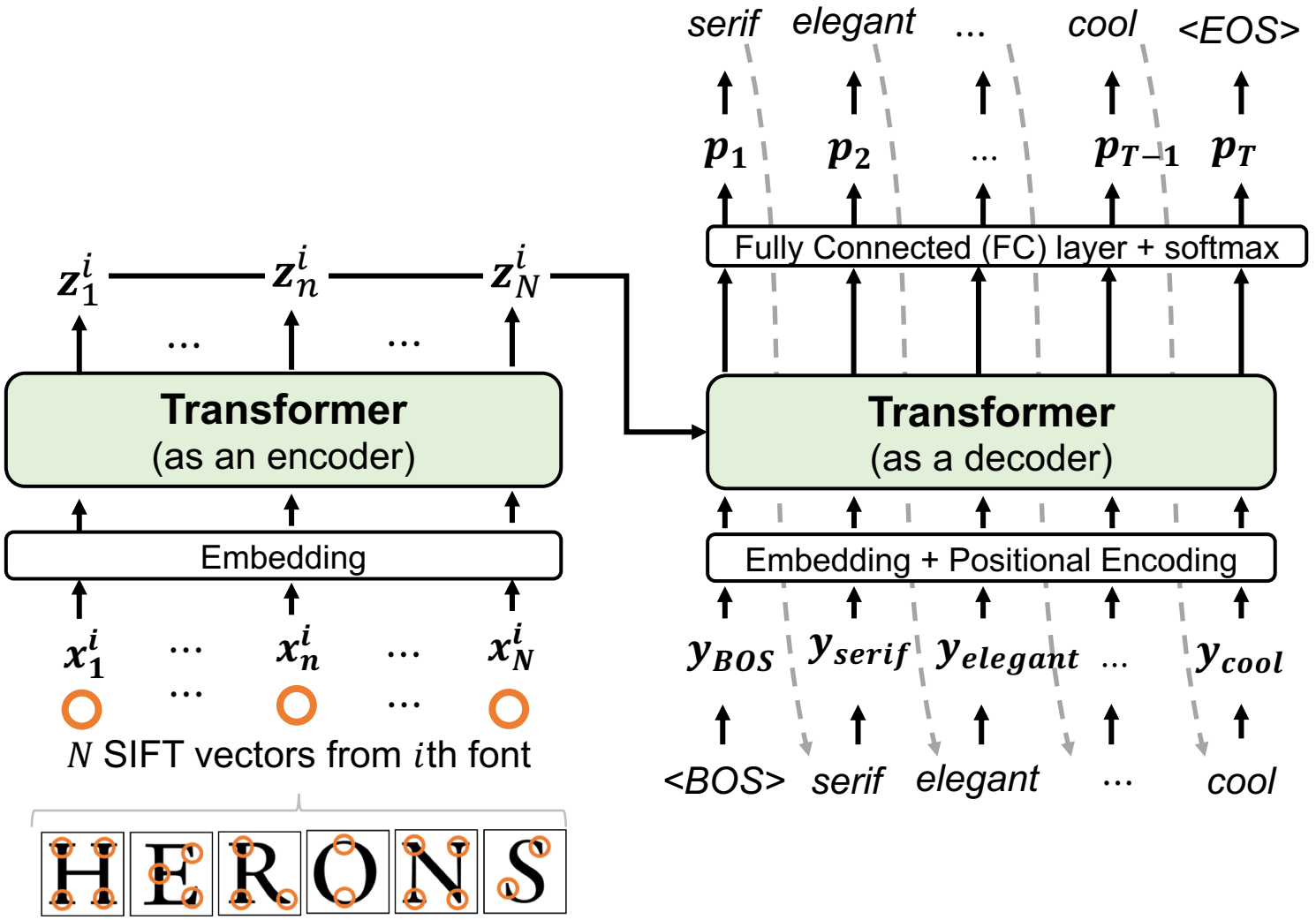


③ Analyze the important parts for specific impressions by **explainable-AI (X-AI) techniques**

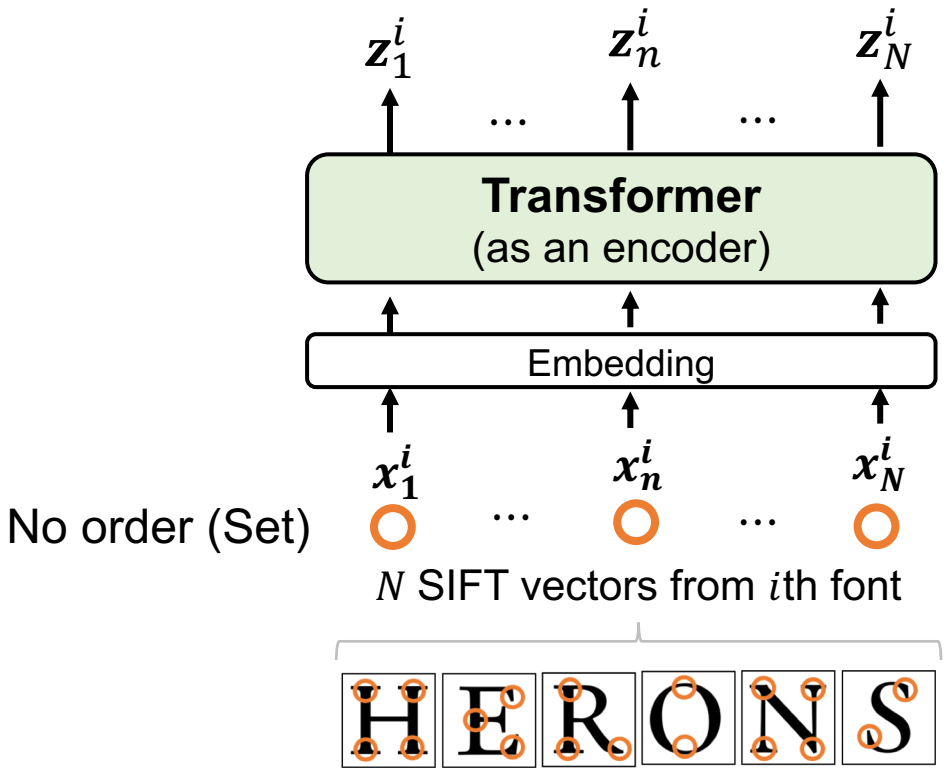
[*] Vaswani et al, ”Attention is all you need”, NeurIPS (2017)

[**] Lowe, D.G., ”Distinctive Image Features from Scale-invariant Keypoints”, Int. J.Comp. Vis.60(2), 91–110 (2004)

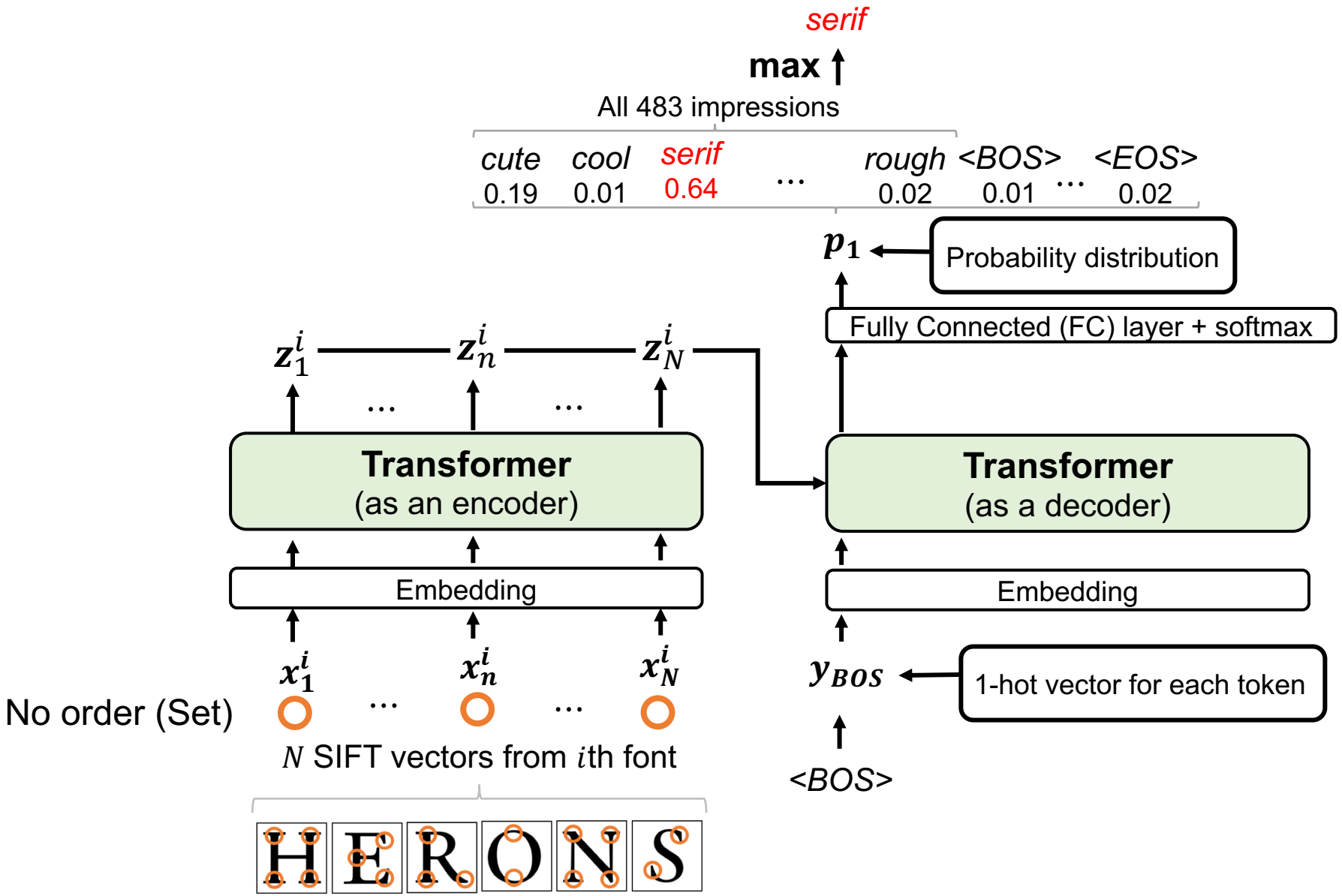
Our Model : Shape-to-Impression Translator



Our Model : Shape-to-Impression Translator



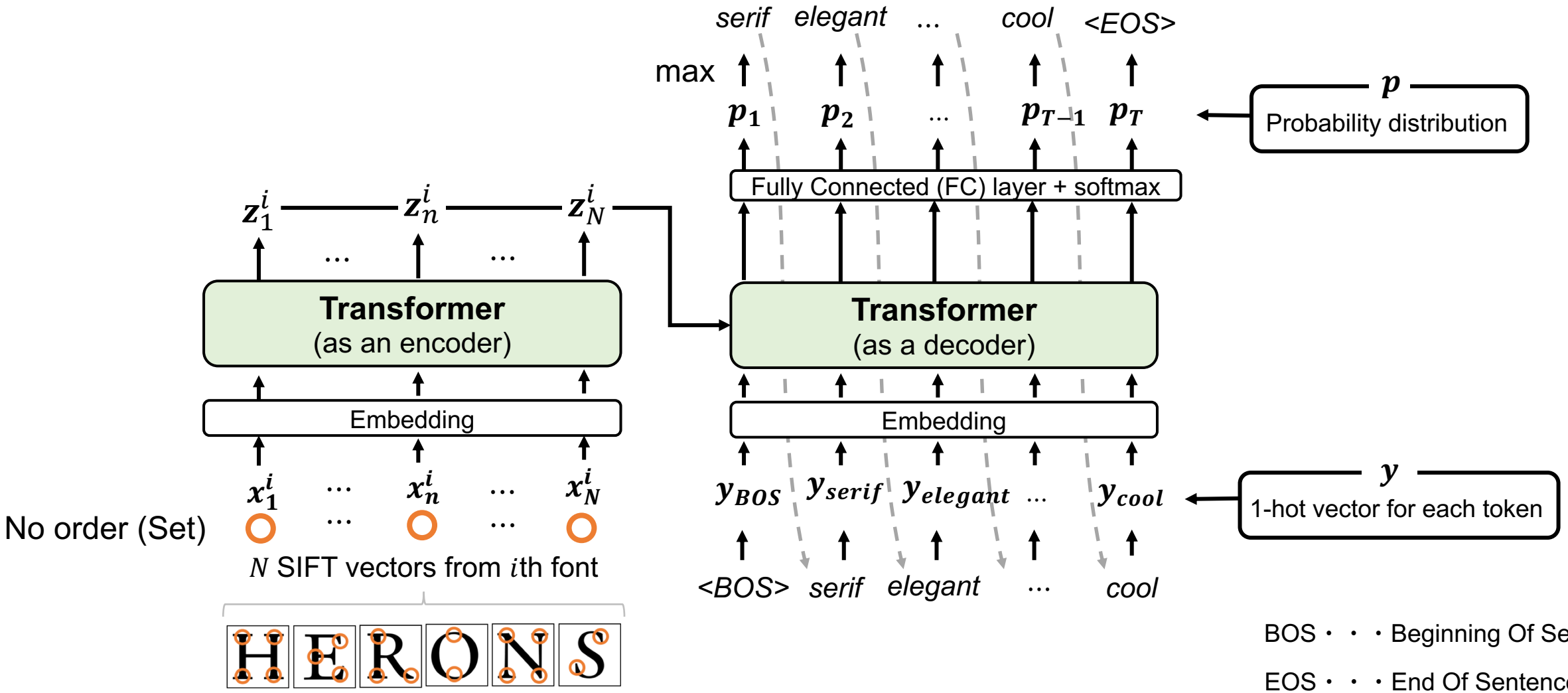
Our Model : Shape-to-Impression Translator



BOS・・・Beginning Of Sentence

EOS・・・End Of Sentence

Our Model : Shape-to-Impression Translator



BOS・・・Beginning Of Sentence

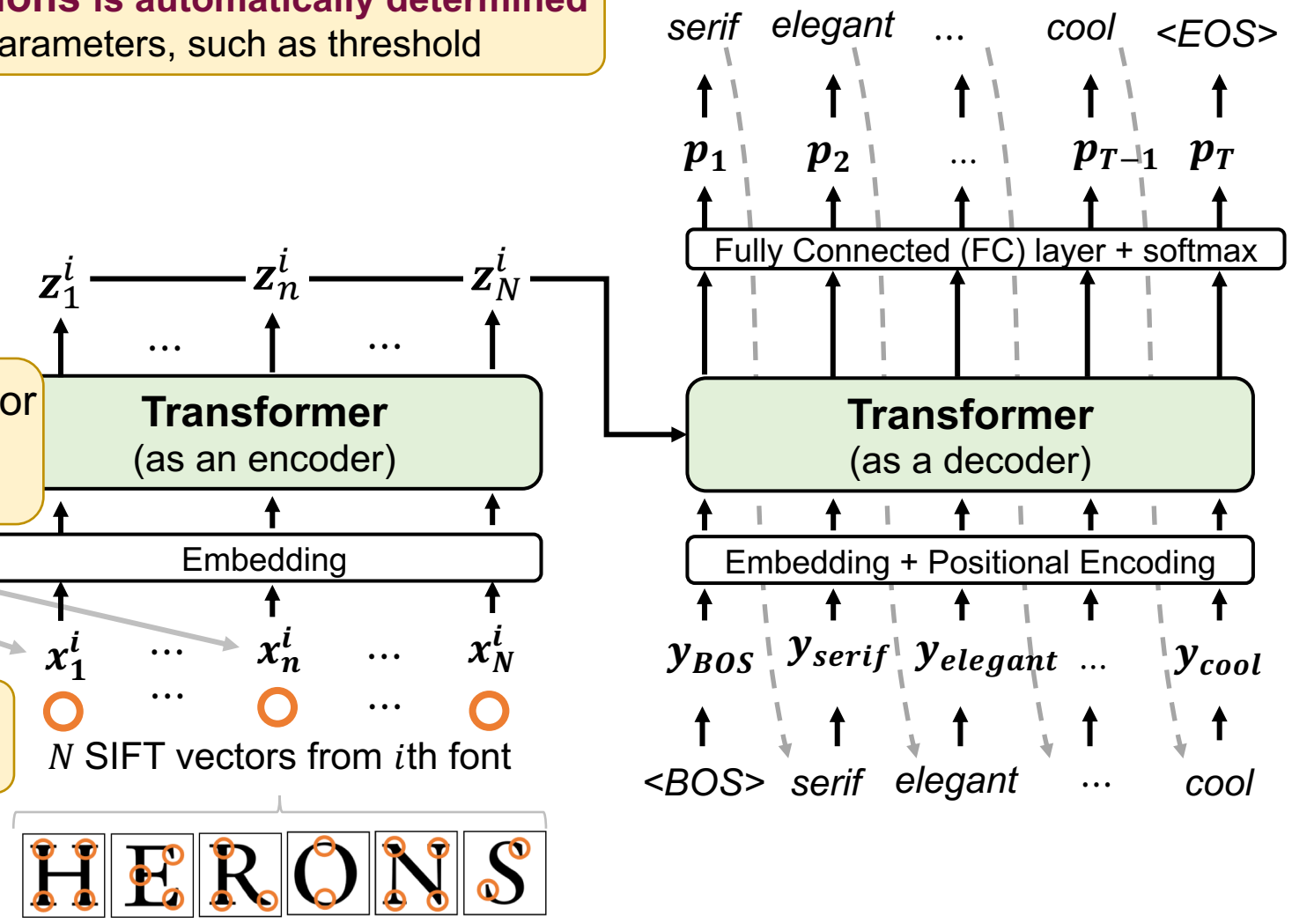
EOS・・・End Of Sentence

Advantages of Translation Framework

Output impressions is automatically determined
without any hyperparameters, such as threshold

Self-attention determines weights for every vector
- **Boost the correlation** among the local shapes that are important for estimation

Accept **a variable number** of SIFT vectors
- The number of local shapes(***N***) **is variable** per font



Experimental result (1/2)

Accuracy of the translation results

Font-Impression Dataset [*]

- **18,815 fonts** and **1,824 impressions** dataset collected from MyFonts.com
- Use the impressions contained **in more than 100 fonts** and the fonts labeled with them
- Train : Validation : Test = **14,876 : 1,856 : 1,847** fonts

18,579 fonts

483 impressions

serif, elegant, narrow, decorative, sensitive, legible, feminine, magazine, business-text, neutral

elegant, swash, script

angle, bold, headline, oblique, retro, italic, display, sans-serif

[*] Chen T. et al., ”Large-scale Tag-based Font Retrieval with Generative Feature Learning”, ICCV(2019)

Translation Examples

Clothe

Original image

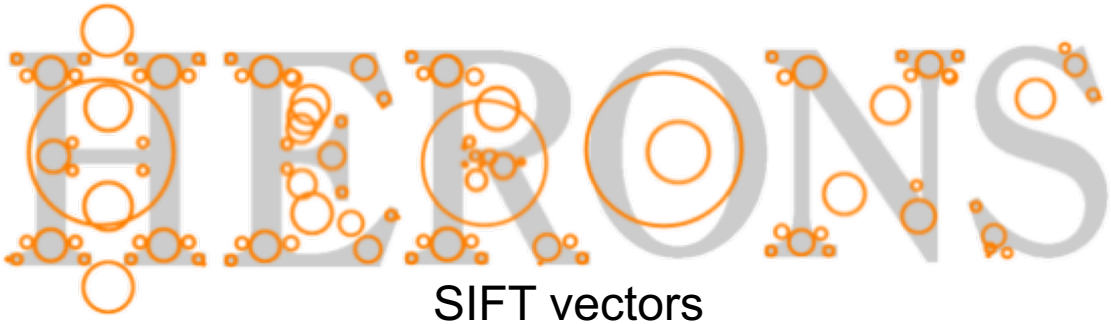
SIFT vectors

Ground Truth

decorative, display, headline, poster, magazine, ***handwrite****, retro,* ***script****,* ***elegant****, informal, logo, vintage, alternate, brand,* ***round****, ligature, invitation, advertise,* ***calligraphy****,* ***handmade****, hand,* ***hand-drawn****,* ***brush****,* ***calligraphic****, cursive, signage, connect, logotype,* ***soft****, contextual-alternates*

Century-old-style-sh

Original image



SIFT vectors

Ground Truth

magazine, ***legible****,* ***serif****,* ***news****, technical, american, masculine, 1900s, transitional,* ***news-headline****,* ***news-text***

Translation Examples

Clothe

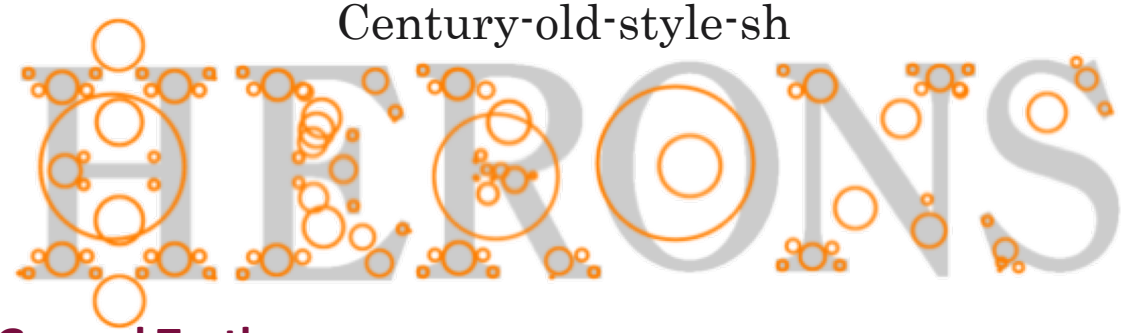
Ground Truth

*decorative, display, headline, poster, magazine, **handwrite**, retro, **script**, **elegant**, informal, logo, vintage, alternate, brand, **round**, ligature, invitation, advertise, **calligraphy**, **handmade**, hand, **hand-drawn**, **brush**, **calligraphic**, cursive, signage, connect, logotype, **soft**, contextual-alternates*

Translation Result

*decorative, display, headline, poster, magazine, **handwrite**, retro, **script**, **elegant**, informal, logo, vintage, alternate, brand, **round**, ligature, invitation, advertise, **calligraphy**, **handmade**, hand, **hand-drawn**, **brush**, **calligraphic**, cursive, signage, connect, logotype, **soft**, contextual-alternates*

Century-old-style-sh



Ground Truth

magazine**, **legible**, **serif**, **news**, technical, american, masculine, 1900s, **transitional**, news-headline, **news-text

Translation Result

magazine**, **legible**, **serif**, text, book, **news**, newspaper, workhorse, **transitional**, book-text, **news-text

Translation Examples (Common impression of GT and result is *green highlights*)

Clothe

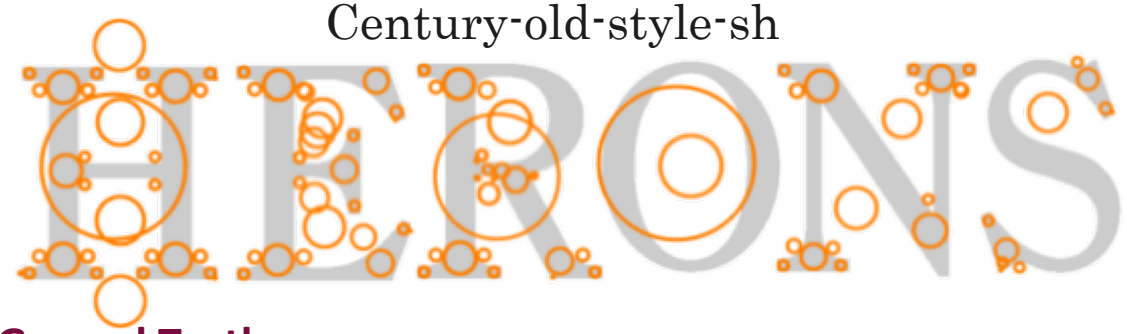
Ground Truth

decorative, display, headline, poster, magazine, ***handwrite****, retro,* ***script****,* ***elegant****, informal, logo, vintage, alternate, brand,* ***round****, ligature, invitation, advertise,* ***calligraphy****,* ***handmade****, hand,* ***hand-drawn****,* ***brush****,* ***calligraphic****, cursive, signage, connect, logotype,* ***soft****, contextual-alternates*

Translation Result

decorative, display, headline, poster, magazine, ***handwrite****, retro,* ***script****,* ***elegant****, informal, logo, vintage, alternate, brand,* ***round****, ligature, invitation, advertise,* ***calligraphy****,* ***handmade****, hand,* ***hand-drawn****,* ***brush****,* ***calligraphic****, cursive, signage, connect, logotype,* ***soft****, contextual-alternates*

Century-old-style-sh



Ground Truth

magazine*,* ***legible****,* ***serif****,* ***news****, technical, american, masculine, 1900s,* ***transitional****, news-headline,* ***news-text***

Translation Result

magazine*,* ***legible****,* ***serif****, text, book,* ***news****, newspaper, workhorse,* ***transitional****, book-text,* ***news-text***

Translation Examples (Common impression of GT and result is *green highlights*)

Clothe

Ground Truth

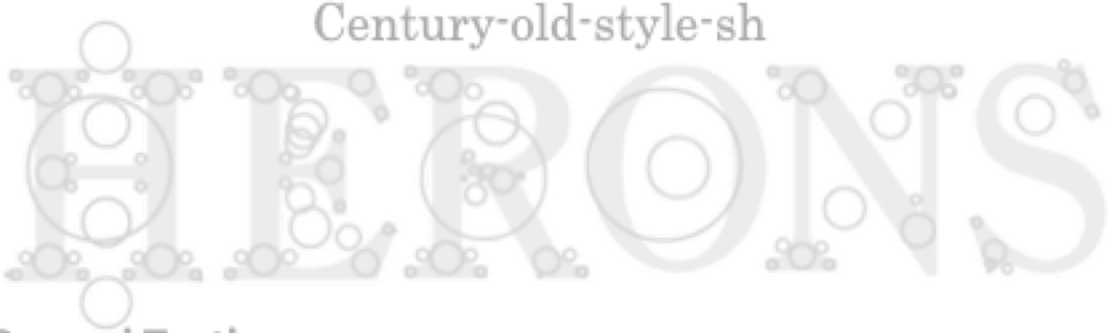
decorative, display, headline, poster, magazine, ***handwrite****, retro,* ***script****,* ***elegant****, informal, logo, vintage, alternate, brand,* ***round****, ligature, invitation, advertise,* ***calligraphy****,* ***handmade****, hand,* ***hand-drawn****,* ***brush****,* ***calligraphic****, cursive, signage, connect, logotype,* ***soft****, contextual-alternates*

Translation Result

decorative, display, headline, poster, magazine, ***handwrite****, retro,* ***script****,* ***elegant****, informal, logo, vintage, alternate, brand,* ***round****, ligature, invitation, advertise,* ***calligraphy****,* ***handmade****, hand,* ***hand-drawn****,* ***brush****,* ***calligraphic****, cursive, signage, connect, logotype,* ***soft****, contextual-alternates*

Prediction Perfectly !

Century-old-style-sh



Ground Truth

magazine, legible, serif, news, technical, american, masculine, 1900s, transitional, news-headline, news-text

Translation Result

magazine, legible, serif, text, book, news, newspaper, workhorse, transitional, book-text, news-text

Translation Examples (Common impression of GT and result is *green highlights*)

Clothe

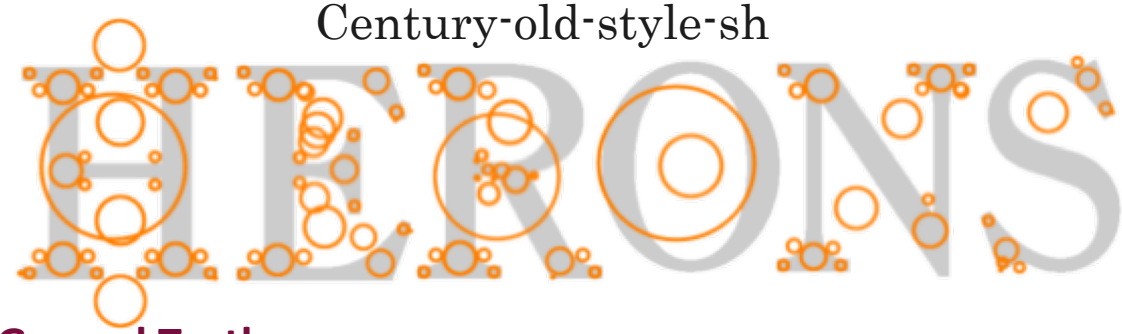
Ground Truth

decorative, display, headline, poster, magazine, handwrite, retro, script, elegant, informal, logo, vintage, alternate, brand, round, ligature, invitation, advertise, calligraphy, handmade, hand, hand-drawn, brush, calligraphic, cursive, signage, connect, logotype, soft, contextual-alternates

Translation Result

decorative, display, headline, poster, magazine, handwrite, retro, script, elegant, informal, logo, vintage, alternate, brand, round, ligature, invitation, advertise, calligraphy, handmade, hand, hand-drawn, brush, calligraphic, cursive, signage, connect, logotype, soft, contextual-alternates

Century-old-style-sh



Ground Truth

magazine, legible, serif, news, technical, american, masculine, 1900s, transitional, news-headline, news-text

Translation Result

magazine, legible, serif, text, book, news, newspaper, workhorse, transitional, book-text, news-text

text, book, book-text, newspaper

- They are similar to GT impressions
- They might be included in the ground-truth

Experimental result (2/2)

Font Shape-Impression Relationship Analysis

Contribution Visualization of Inputs by Integrated Gradients [*]

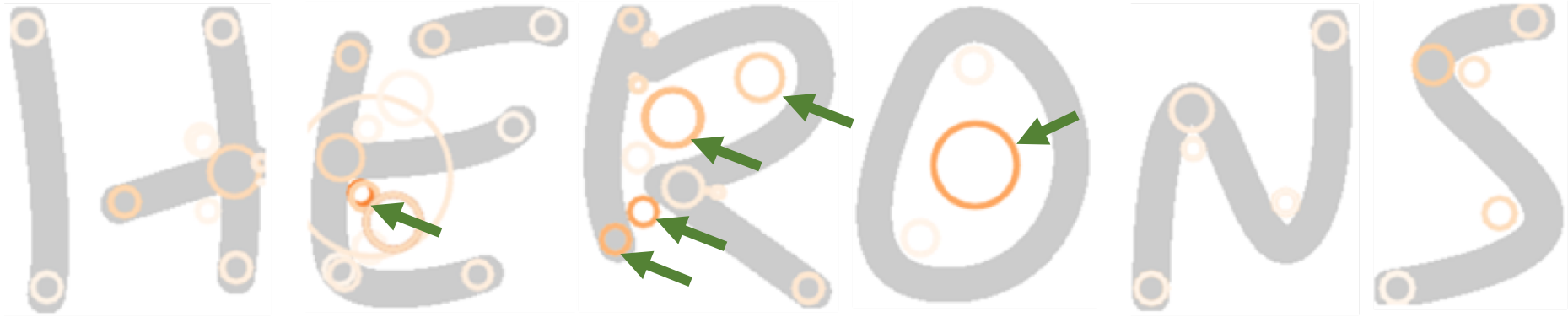
Thicker orange color shape are more important

The shapes of large space have large contribution for *comic-text*

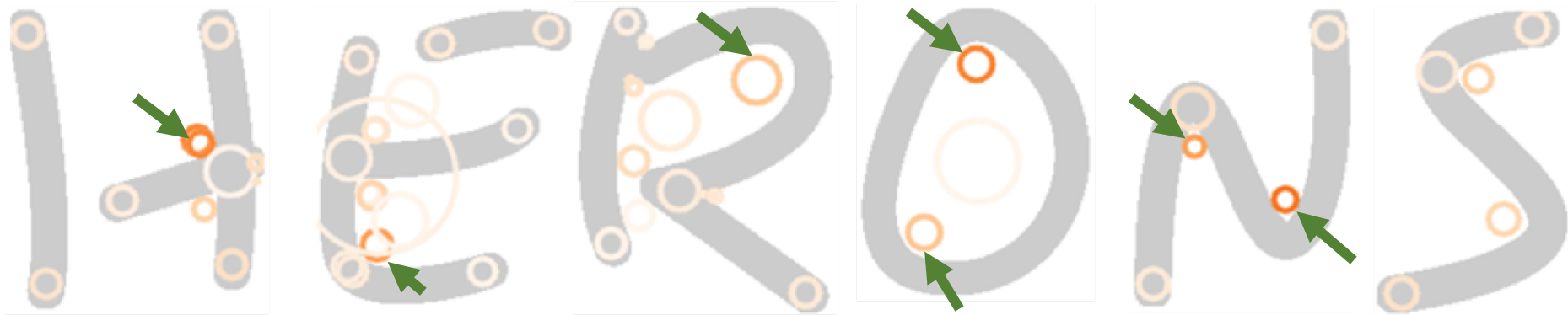
Ales-script

Translation Result

comic-text



script



The shapes of sharp corner have large contribution for *script*

[*] Sundararajan, M et al, ” Axiomatic Attribution for Deep Networks”, ICLR, 2017

Contribution Visualization of Inputs by Integrated Gradients [*]

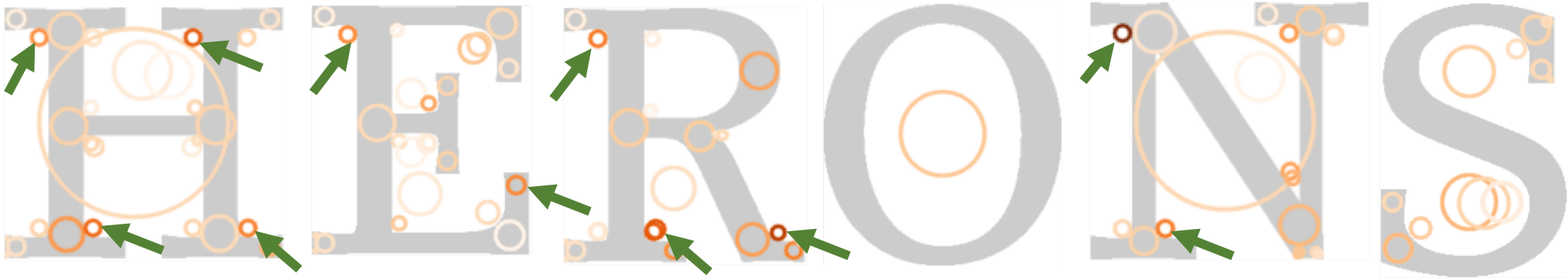
Thicker orange color shape are more important

The shapes of serif-part have large contribution for *comic-text*

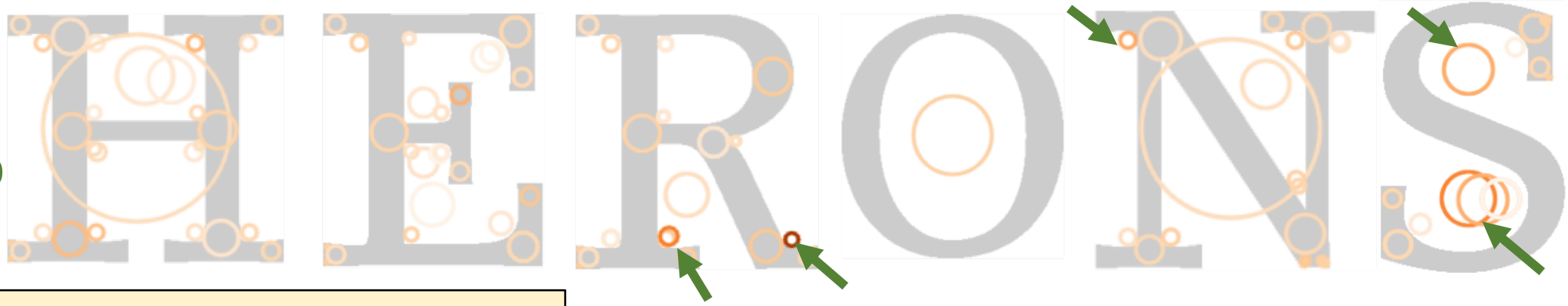
Linoletter

Translation Result

text



legible



The shapes of rounded space have large contribution for *script*

[*] Sundararajan, M et al, ” Axiomatic Attribution for Deep Networks”, ICLR, 2017

Concluding remarks

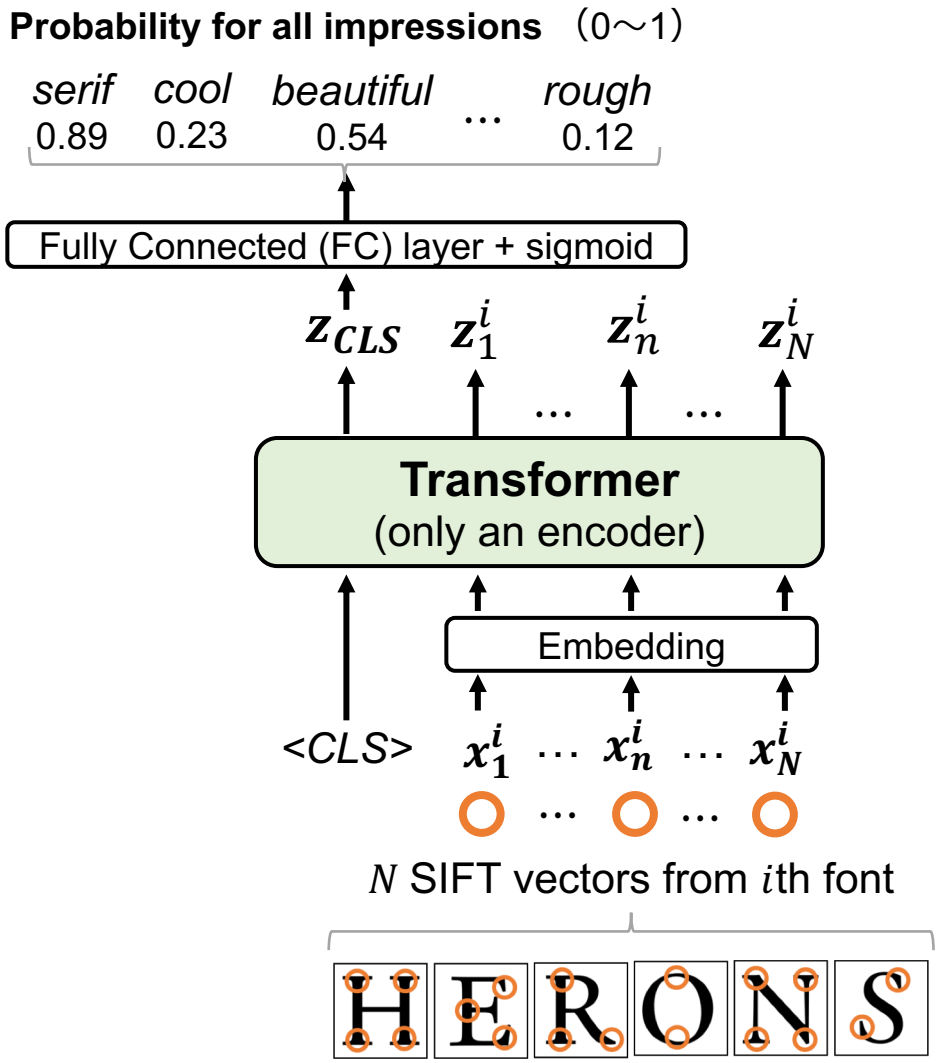
Conclusion

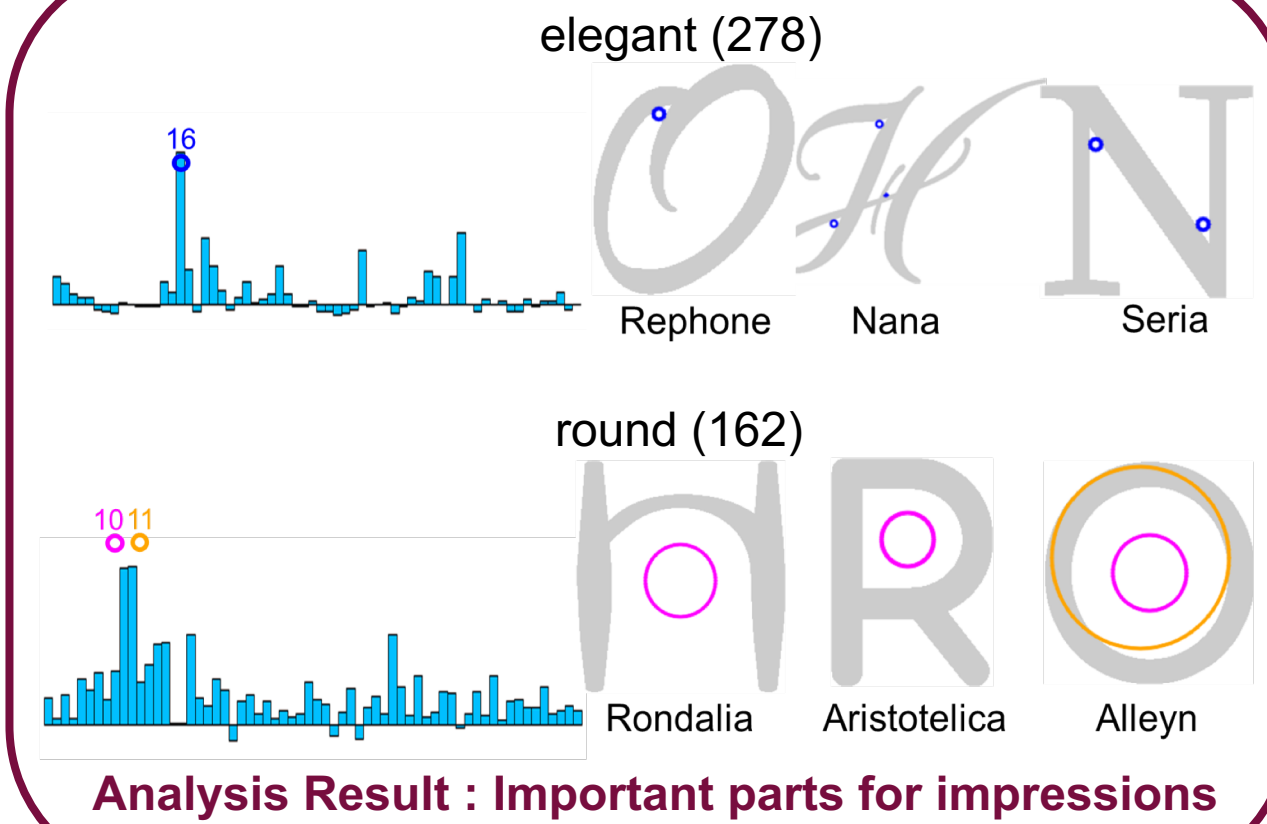
- Propose the part-based font-impression analysis using Transformer
- Confirm that Translator can correctly output impressions from a set of local shapes
- Analyze important local shapes for specific impressions by using the trained Transformer with explainable-AI techniques

Future Work

- Practical applications of the proposed methods
- Collaboration with experts of cognitive psychology

Another Simpler Model : Multi-label Classifier

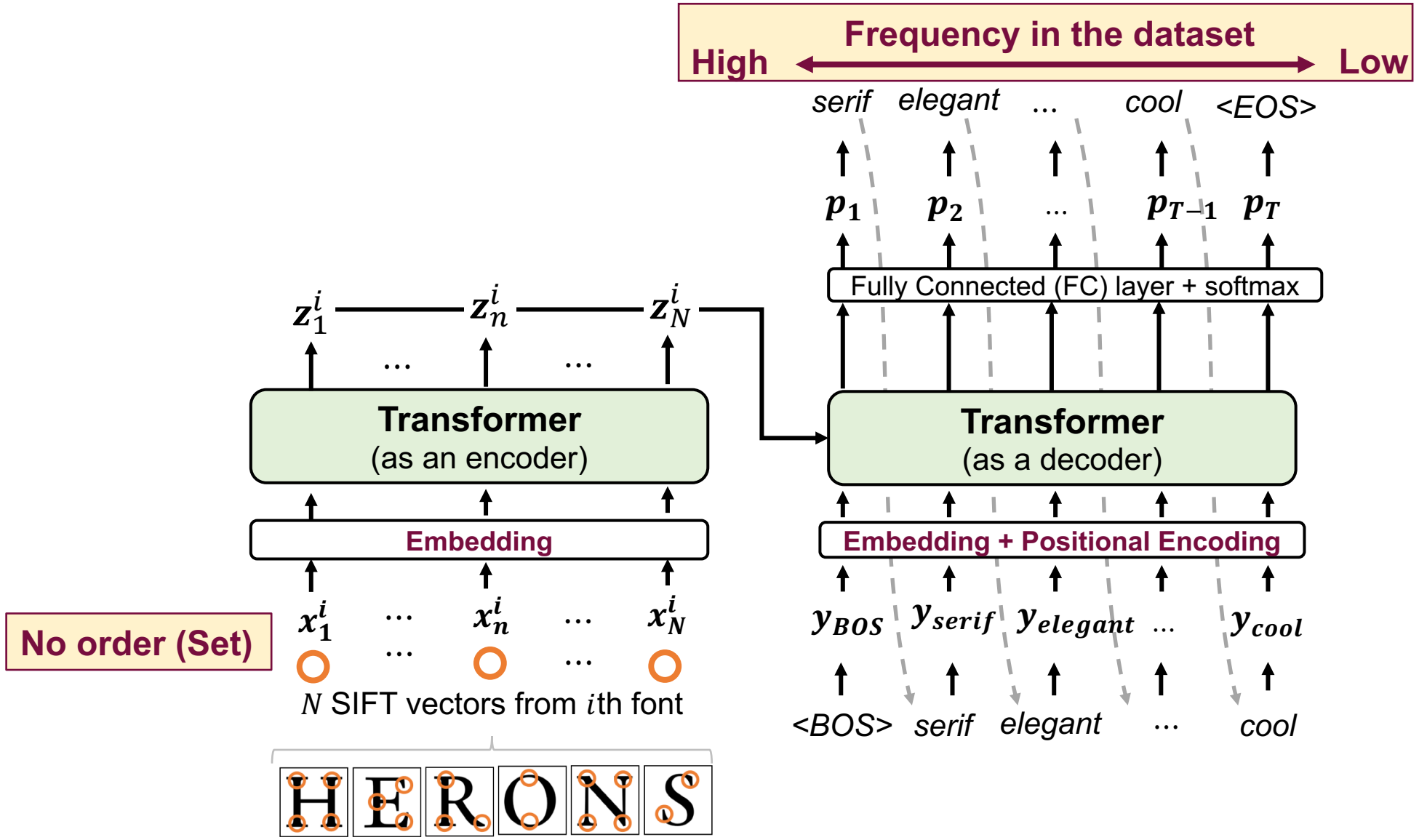




For its details,
please refer to our paper

ここまで

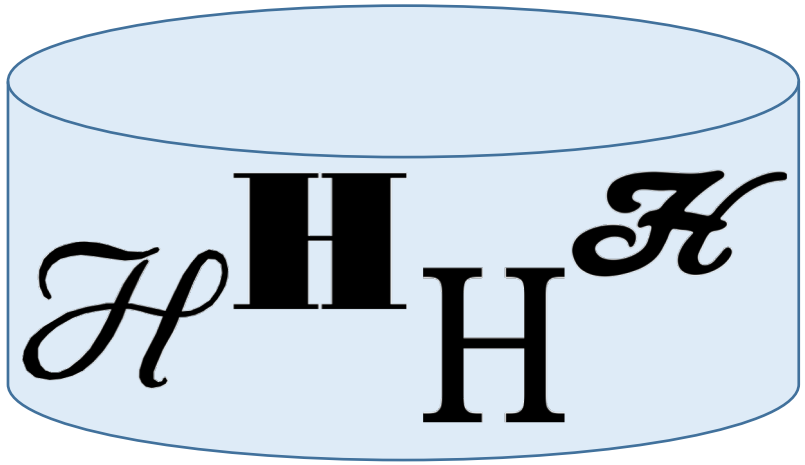
The Order of Inputs and Outputs



Applications

Font recommendation system

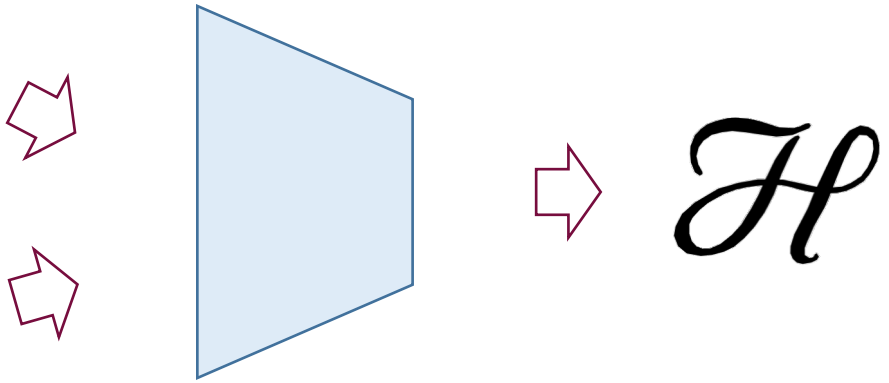
Please *handwrite* font



Font generation system

Important parts for *handwrite*

H



Why Did We Use SIFT?

- SIFT is so convenience because it gives us more explainability
- There is no difference between SIFT and SURF

Quantitative Evaluation by **Multi-label Classifier**

	SIFT	SURF
F1@all	**0.145**	0.140
mAP	0.135	**0.151**

Strong and Weak Impression Words

Comparison among Methods

- **Multi-label classifier** outperforms the other models on all measures
- **Translator** does not require a threshold

	Multi-label Classifier	Translator	ViT [*]	DeepSets [**]
Inputs	**SIFT**	SIFT	Patch	SIFT
F1@100↑	**0.301**	0.264	0.264	0.279
F1@200↑	**0.221**	0.186	0.185	0.194
F1@all↑	**0.145**	0.117	0.109	0.110
mAP↑	**0.135**	N/A	0.115	0.115
Threshold	✓		✓	✓

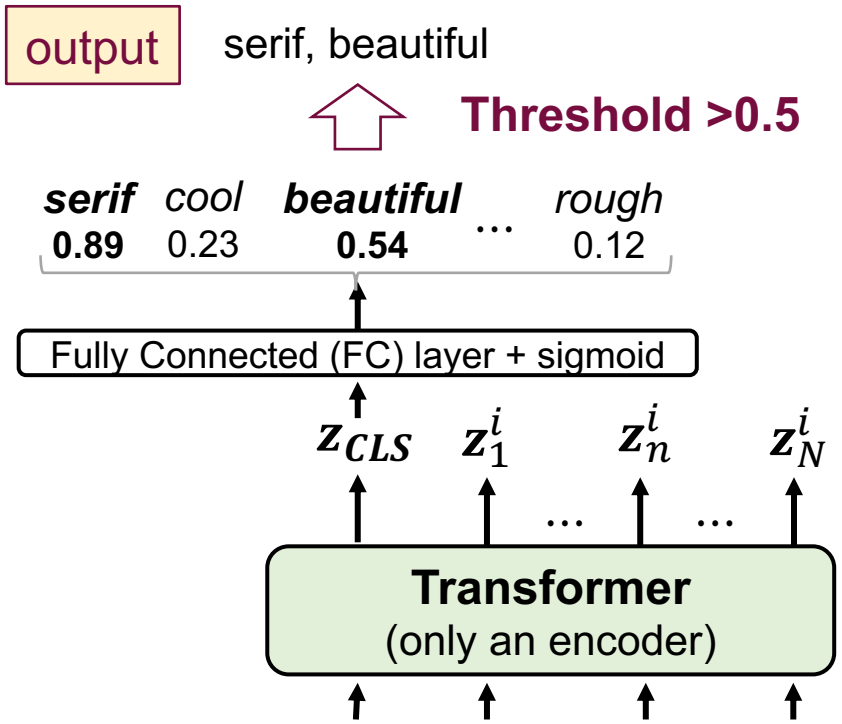
F1@n ・・・F1 score with frequency top n impressions in the dataset

[*] Dosovitskiy, A et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale” , ICLR(2021)
[**] Ueda, M, et al., ”Which Parts Determine the Impression of the Fonts?”, ICDAR (2021)

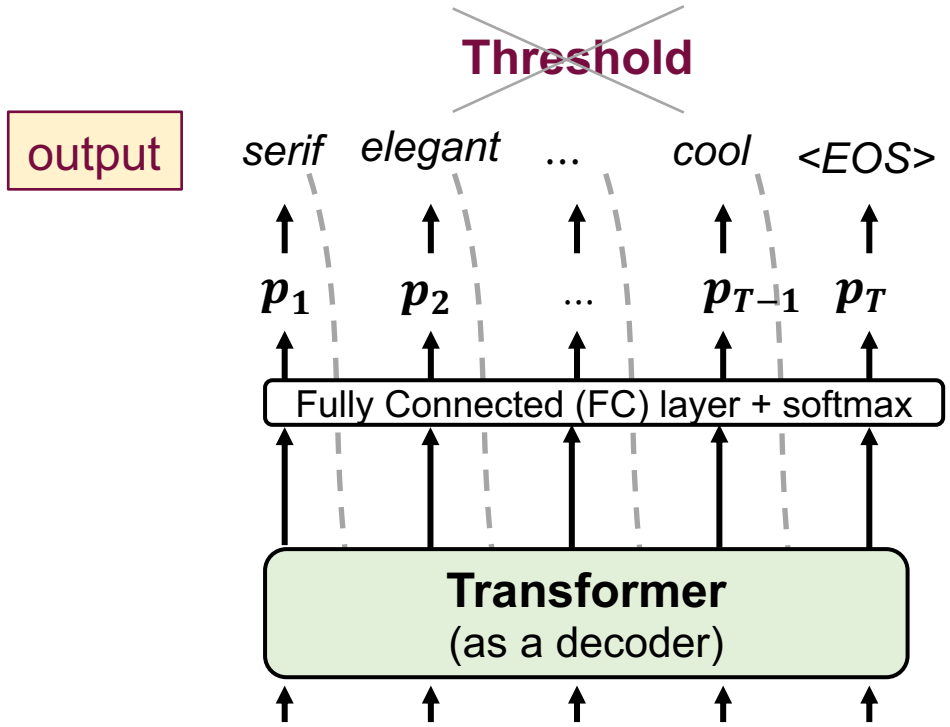
Translator Does Not Require a Hyperparameter

EOS token **determines the number** of output impressions

- In multi-label classifier, we **have to determine hyperparameter (threshold)**



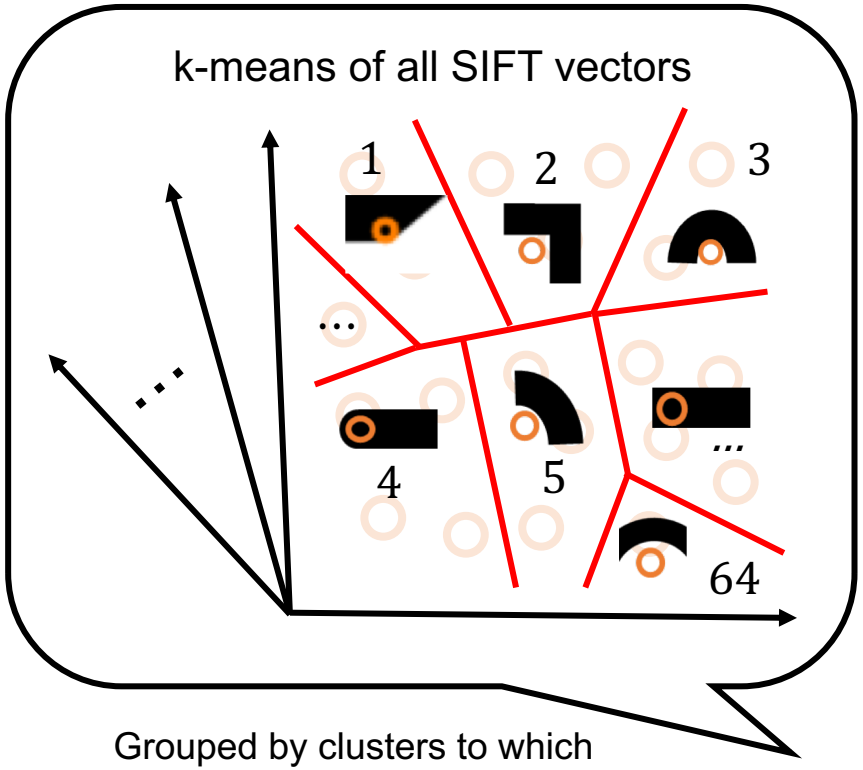
Multi-label Classifier



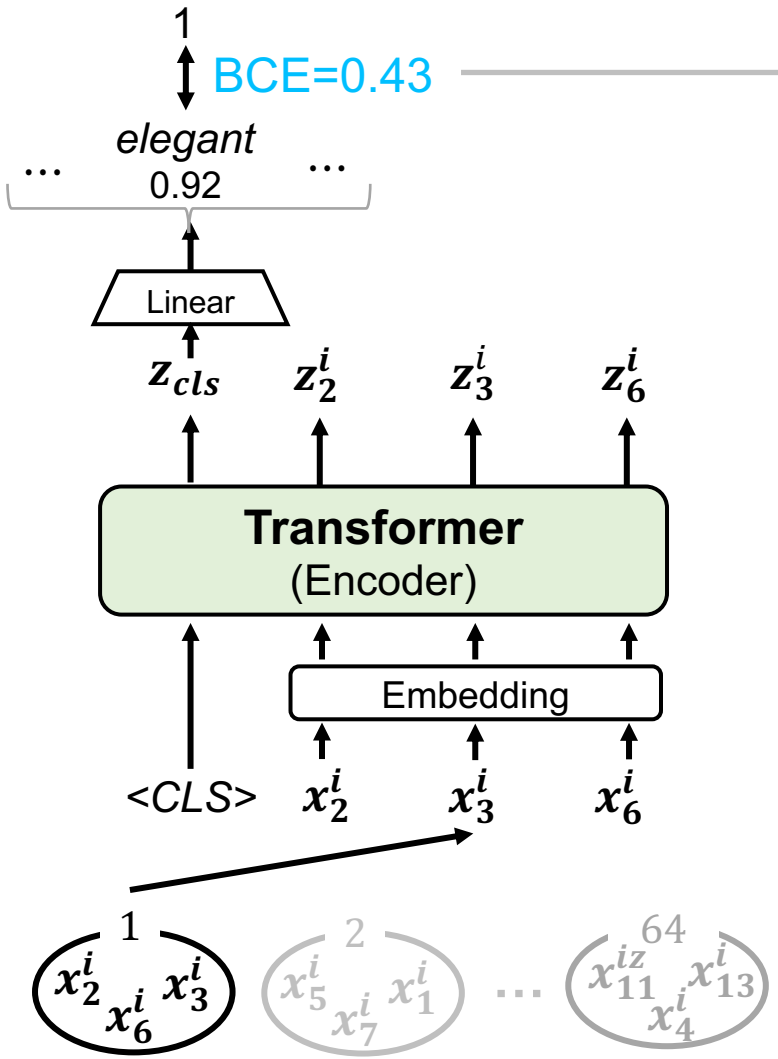
Translator

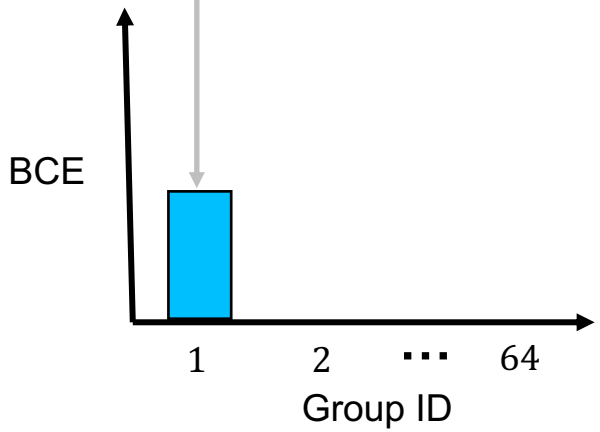
Group-Based Occlusion Sensitivity

Example : *elegant*



Grouped by clusters to which the SIFT vector belongs



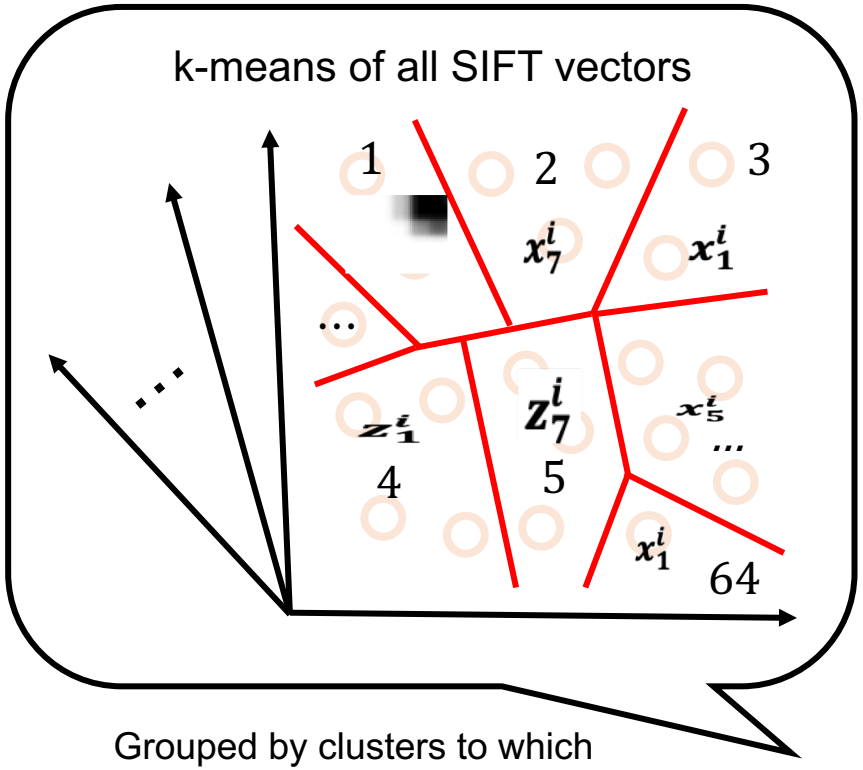


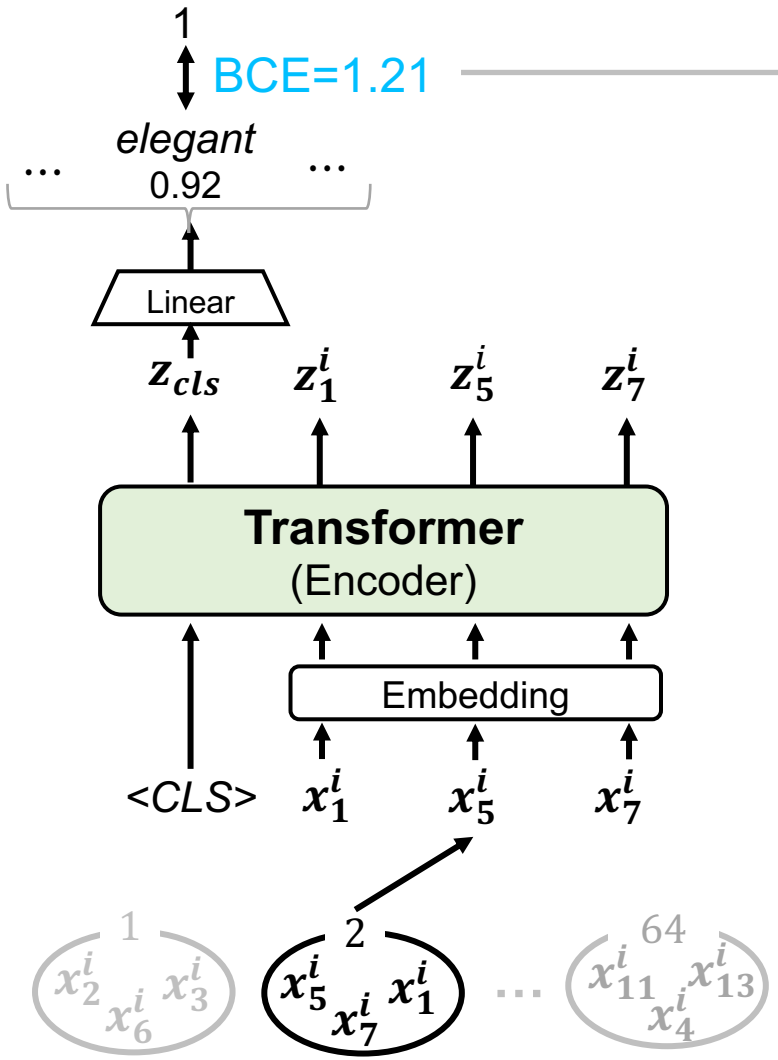
Contribution to *elegant* by shape

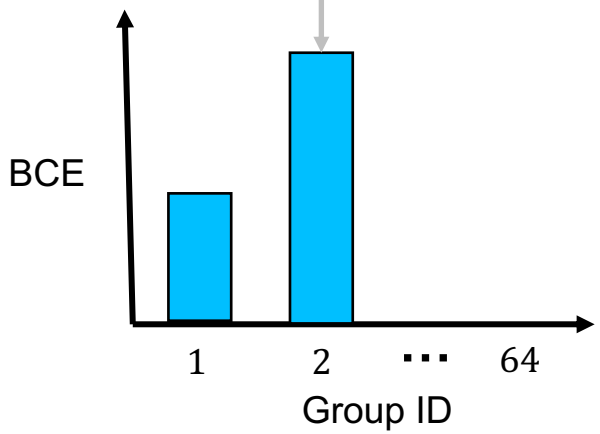
HERONS

Group-Based Occlusion Sensitivity

Example : *elegant*





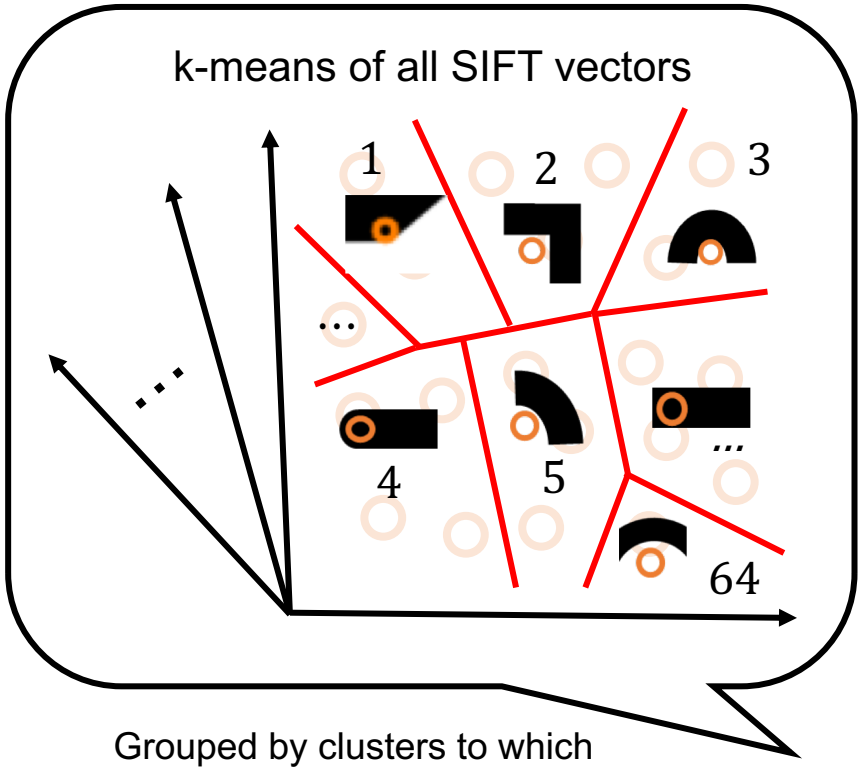


Contribution to *elegant* by shape

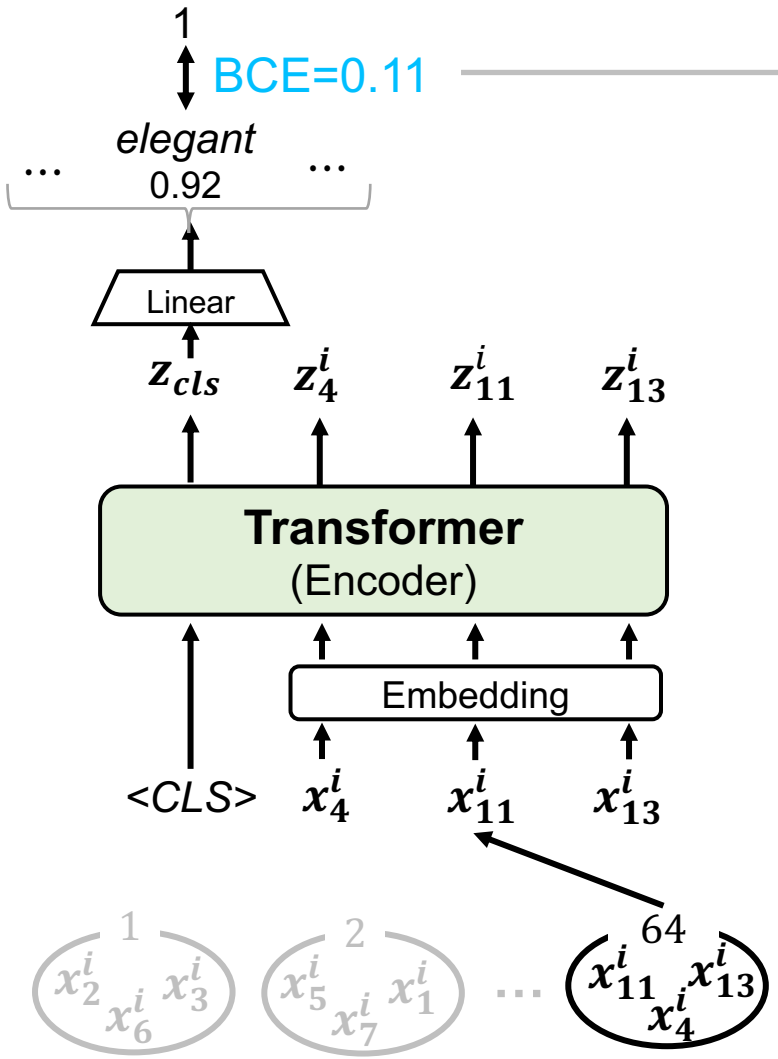
HERONS

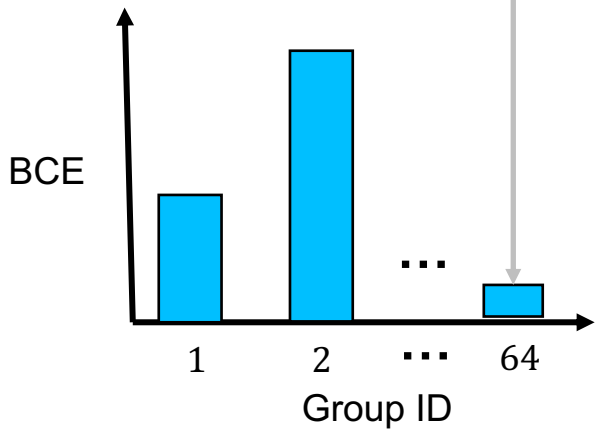
Group-Based Occlusion Sensitivity

Example : *elegant*



Grouped by clusters to which the SIFT vector belongs





Contribution to *elegant* by shape

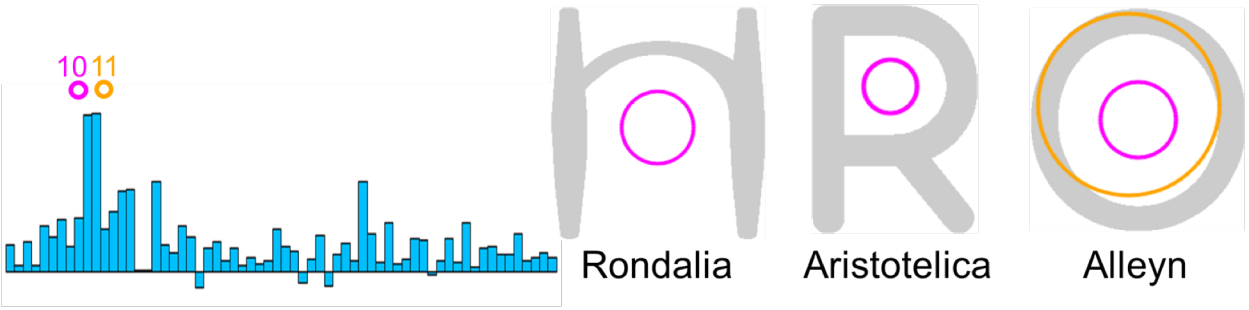
HERONS

Important Local Shapes for Specific Impressions

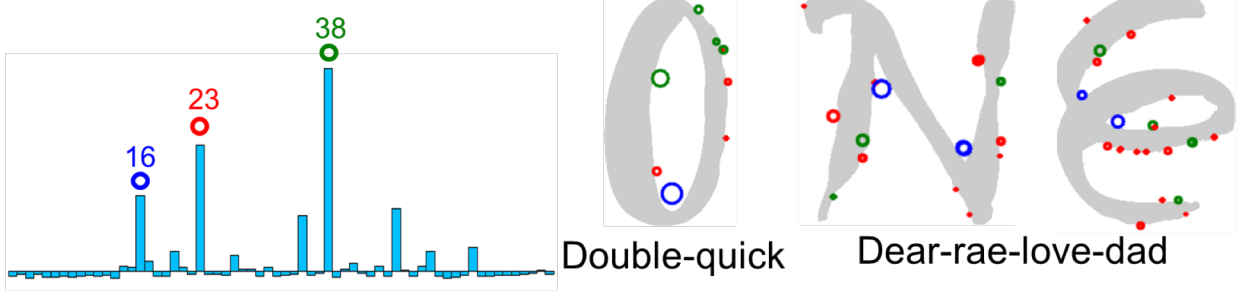
Create weighted histogram of the shapes that are important for based on **specific impressions**

the number of test data

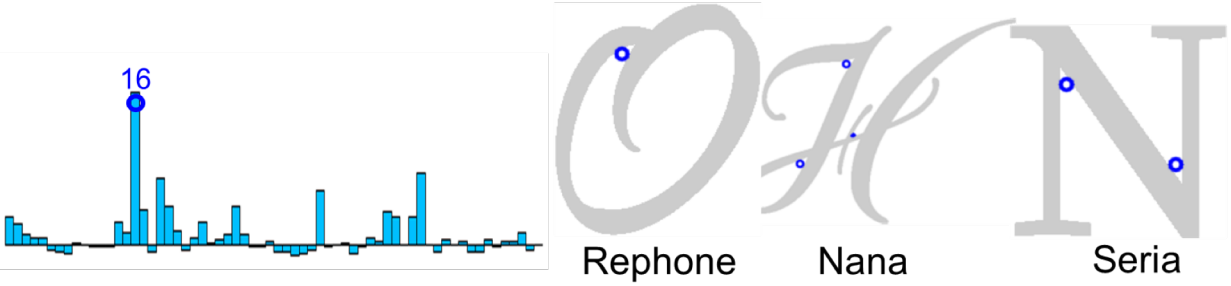
round (162)



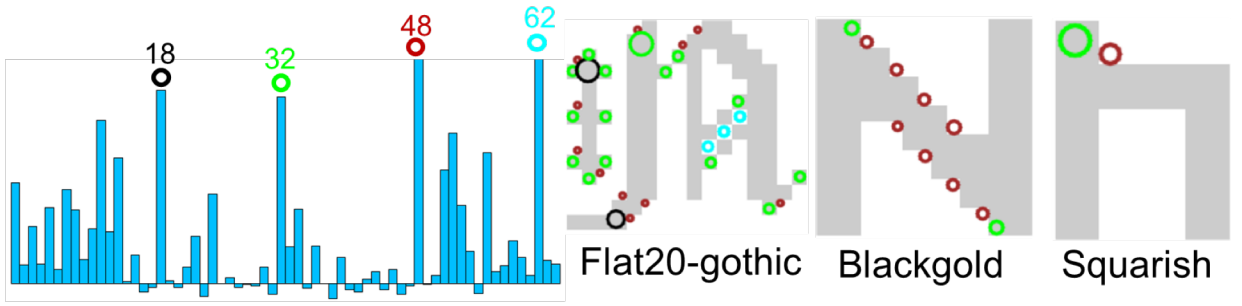
handwrite (321)



elegant (278)



computer (58)



Contribution Visualization of Inputs by Integrated Gradients [*]

$$Integrated\ Gradients(IG) = (\boldsymbol{x} - \boldsymbol{x}') \int_0^1 \nabla F\left(\boldsymbol{x}' + \alpha(\boldsymbol{x} - \boldsymbol{x}')\right) d\alpha$$

$\boldsymbol{x} \in \boldsymbol{R^d}$: input vector（SIFT vector）

$\boldsymbol{x}' \in \boldsymbol{R^d}$: baseline (0 vector)

$F : \boldsymbol{R}^d \rightarrow$ [0, 1] that represents a deep network (Transformer)

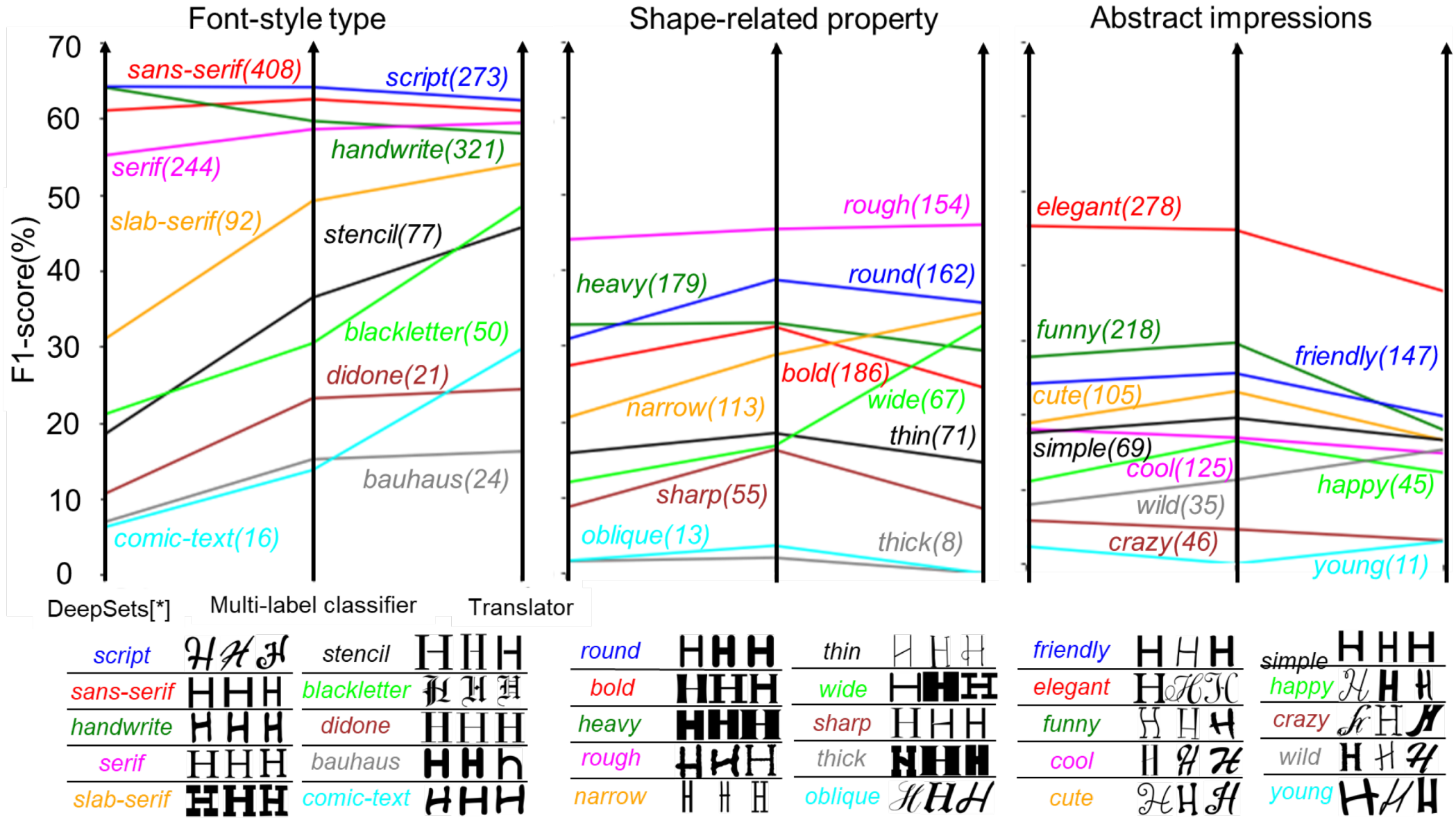
Contribution Map (by IG)

Translation Result

	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	…	$\boldsymbol{x_N}$
serif	0.32	0.12	…	0.15
elegant	-0.23	0.4	…	0.02
script	-0.11	-0.11	…	0.32
<EOS>	0.01	-0.24	…	-0.12

[*] Sundararajan, M et al, ” Axiomatic Attribution for Deep Networks”, ICLR, 2017

Comparison of F1-score by 3 methods



[*] Zaheer, M, et al.,"Deep Sets", NeurIPS (2017)

Metrics : F1-score

- Binary classification is performed for each impression, and the F1-score is calculated for each impression.
- Average F1-score for each impression

	Font1	Font2	Font3	Font4
Label	[0, 1, 1, 1]	[1, 0, 1, 0]	[0, 1, 1, 0]	[1, 1, 1, 1]
Estimation	[1, 0, 1, 1]	[1, 1, 1, 1]	[1, 0, 0, 1]	[1, 0, 1, 0]

Impression1

Impression1

	E P	E N
GT P	2	0
GT N	2	0

Impression2

	E P	E N
GT P	0	3
GT N	1	0

Impression3

	E P	E N
GT P	3	1
GT N	0	0

Impression$

	E P	E N
GT P	1	1
GT N	2	0

Metrics：Average Precision（AP）

Ranking of all fonts with impression X as label Prediction probability $\{r_1, r_2 \ldots r_H\}$

H is the number of fonts with impression *X* as a label

$$AP = \frac{1}{H}\sum_{h=1}^{H}\frac{h}{r_h}$$

Ex.：decorative

Rank	1	2	3	4	5	6	· · ·	n
Probability	0.99	0.91	0.83	0.74	0.71	0.7		0.001
Label or Not	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE		FALSE

$$AP_{decorative} = \frac{1}{Sum\ of\ True}\left(\frac{1}{1} + \frac{2}{3} + \frac{3}{6} + \cdots\right)$$

Contribution Visualization of Inputs by Integrated Gradients [*]

Thicker orange color shape are more important

The shapes of serif-part are deeply visualized

Slam-normal

Translation Result

serif

slab-serif

The shapes of right-angle part are deeply visualized

[*] Sundararajan, M et al, ” Axiomatic Attribution for Deep Networks”, ICLR, 2017

Questions

- ~~入出力の順序は？~~
- ~~アプリケーションは？~~
- ~~最後にチラッと言っていたマルチラベル認識の結果は~~
- comparative studyは？
  - そもそも類例が少ないのだが自分たちがやった範囲では
- ~~SIFTじゃないとだめなのか？~~
  - ~~SURFのやつを出す~~
- ~~どれぐらいの性能がでれば使えるのか？~~
- ~~得意な印象語と不得意な印象語は？~~