The 36th annual meeting of Japanese cognitive science society

口頭発表 3 O3-4

All about Attention

浅川伸一 (**あさかわ しんいち**) asakawa@ieee.org

礼和元年 9 **月** 6 **日**

モチベーション

- 2018 年 人間超えした自然言語処理モデル
- 認知科学，神経心理学，認知心理学，生理学，計算モデルでの知見が蓄積
- 上記の関係を考える
- 最後に少しだけ計算例

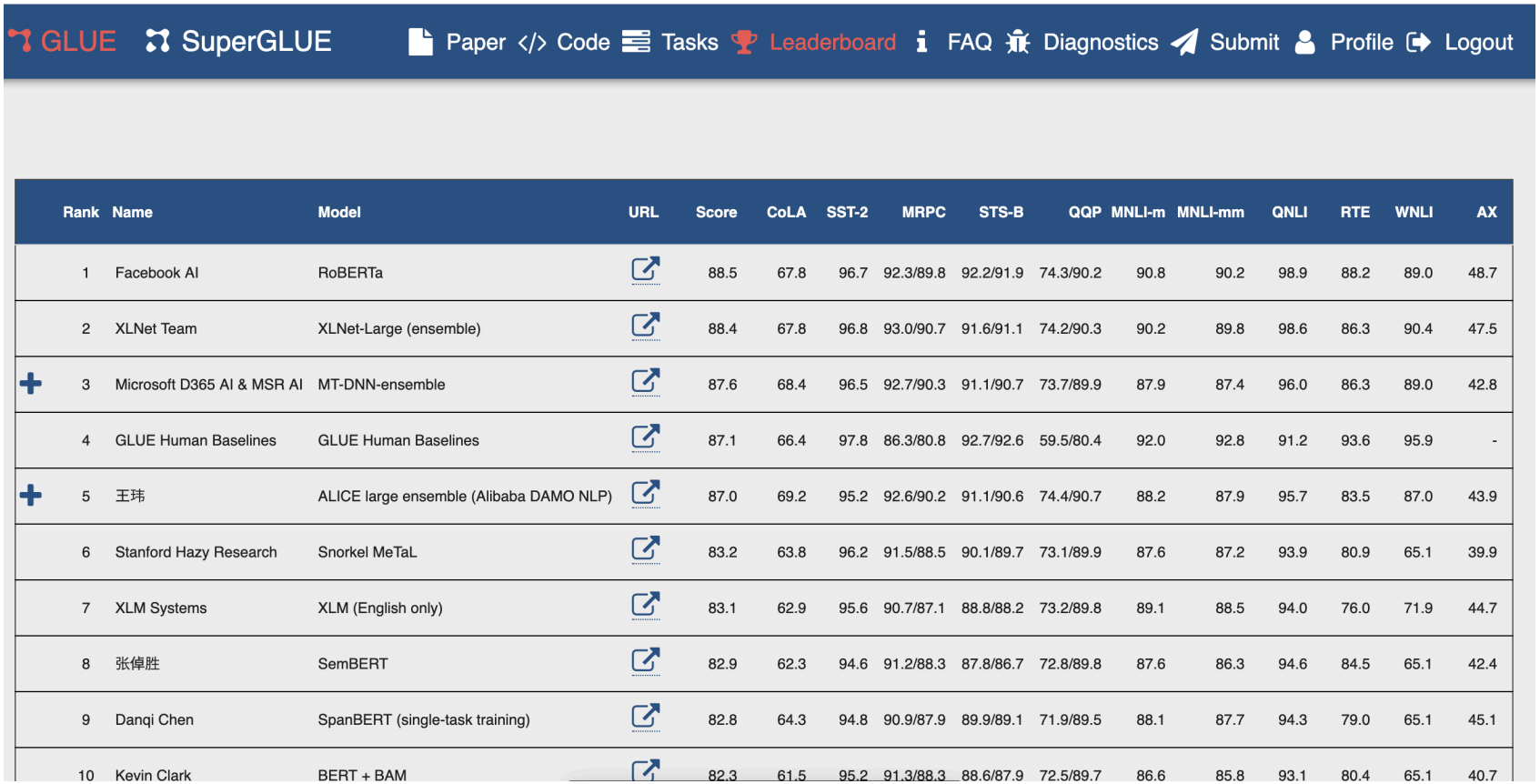
本発表の構成

1. **自然言語処理分野の注意，トランスフォーマー**
2. **認知心理学，生理学，神経心理学の知見，計算モデル**
3. DeepGazeII
4. **計算例と考察**

Takeaways

自然言語処理 (Liu, He, Chen, & Gao, 2019; Liu et al., 2019; Vaswani et al., 2017) **や眼球運動での** SOTA **である** DeepGazeII(Kummerer, Wallis, Gatys, & Bethge, 2017) **ではボトムアップ注意が用いられている。一方，認知心理学，生理学，などではトップダウンとボトムアップの** 2 **種類の注意が区別されてきた。いずれも** Crick (1984) **により提唱された勝者占有回路である。トップダウン注意も考慮したモデルを組み込む必要があるだろう**

GLUE リーダーボード



Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI	AX
1	Facebook AI	RoBERTa		88.5	67.8	96.7	92.3/89.8	92.2/91.9	74.3/90.2	90.8	90.2	98.9	88.2	89.0	48.7
2	XLNet Team	XLNet-Large (ensemble)		88.4	67.8	96.8	93.0/90.7	91.6/91.1	74.2/90.3	90.2	89.8	98.6	86.3	90.4	47.5
3	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9	87.9	87.4	96.0	86.3	89.0	42.8
4	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9	-
5	王玮	ALICE large ensemble (Alibaba DAMO NLP)		87.0	69.2	95.2	92.6/90.2	91.1/90.6	74.4/90.7	88.2	87.9	95.7	83.5	87.0	43.9
6	Stanford Hazy Research	Snorkel MeTaL		83.2	63.8	96.2	91.5/88.5	90.1/89.7	73.1/89.9	87.6	87.2	93.9	80.9	65.1	39.9
7	XLM Systems	XLM (English only)		83.1	62.9	95.6	90.7/87.1	88.8/88.2	73.2/89.8	89.1	88.5	94.0	76.0	71.9	44.7
8	张倬胜	SemBERT		82.9	62.3	94.6	91.2/88.3	87.8/86.7	72.8/89.8	87.6	86.3	94.6	84.5	65.1	42.4
9	Danqi Chen	SpanBERT (single-task training)		82.8	64.3	94.8	90.9/87.9	89.9/89.1	71.9/89.5	88.1	87.7	94.3	79.0	65.1	45.1
10	Kevin Clark	BERT + BAM		82.3	61.5	95.2	91.3/88.3	88.6/87.9	72.5/89.7	86.6	85.8	93.1	80.4	65.1	40.7

Figure 1: https://gluebenchmark.com/leaderboard GLUE: General Language Understanding Evaluation

GLUE 下位課題

CoLA: 入力文が英語として正しいか否かを判定

SST-2: スタンフォード大による映画レビューの極性判断

MRPC: マイクロソフトの言い換えコーパス。2 文 が等しいか否かを判定

STS-B: ニュースの見出し文の類似度を 5 段階で評定

QQP: 2 つの質問文の意味が等価かを判定

MNLI: 2 入力文が意味的に含意，矛盾，中立を判定

QNLI: Q and A

RTE: MNLI に似た 2 つの入力文の含意を判定

WNI: ウィノグラッド会話チャレンジ

SOTA モデルの特徴

RoBERTa: BERT の訓練コーパスを巨大 (173GB) にし，ミニバッチサイズを大きした

XLNet: 順列言語モデル。2 ストリーム注意

MT-DNN: BERT ベース の転移学習に重きをおいたモデル

GPT-2: BERT に基づく。人間超えして 2019 年 2 月時点で炎上騒ぎ

BERT: Transformer に基づく言語モデル。マスク化言語モデル と 次文予測 に基づく 事前訓練，各下流課題を ファインチューニング。事前訓練されたモデルは一般公開済。

ELMo: 双方向 RNN による文埋め込み表現

Transformer: 自己注意に基づく言語モデル。多頭注意，位置符号器.

事前訓練とマルチ課題学習

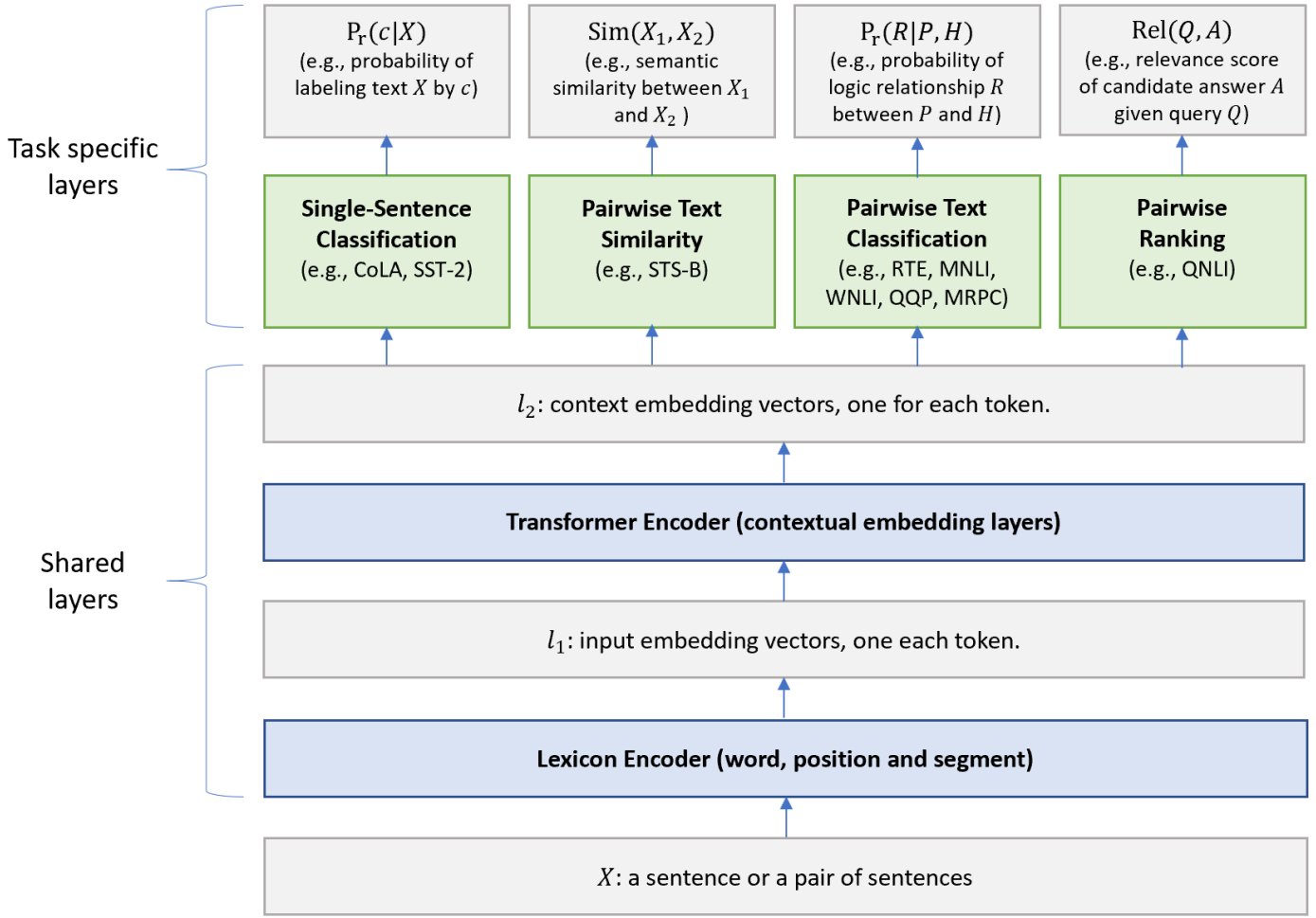


Figure 2: From Liu et al. (2019) Fig. 1

Transformer: Attention is all you need

$$attention\,(Q, K, V) = dropout\left(softmax\left(\frac{QK^{\top}}{\sqrt{d}}\right)\right)V \tag{1}$$

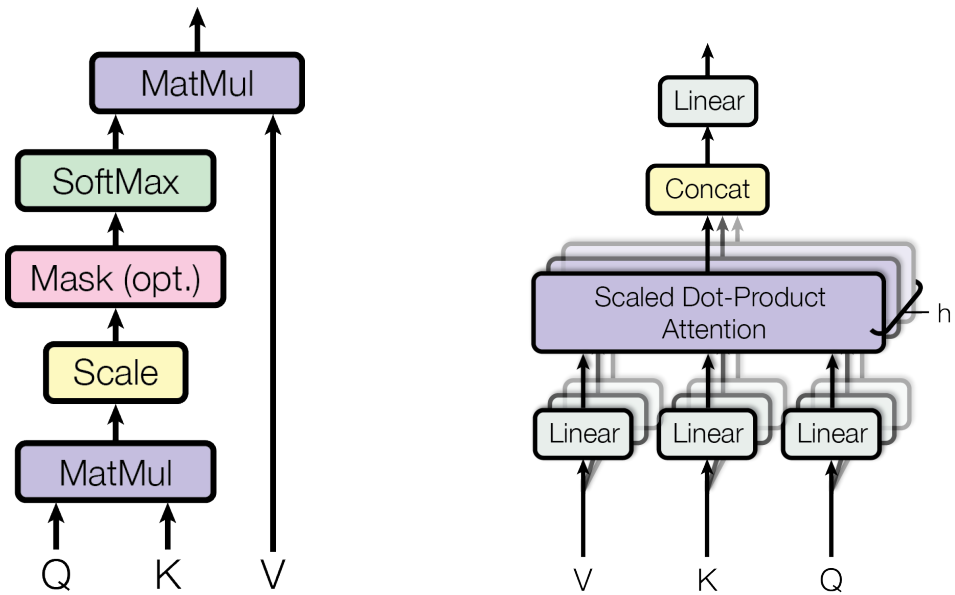


Figure 3: From Vaswani et al. (2017) Fig. 2

Transformer(2): Attention is all you need

$$MultiHead\left(Q,K,V\right)=Concat\left(\underset{1}{head},\ldots,\underset{h}{head}\right)W^{O} \tag{2}$$

where, $head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$

The projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{model}\times d_k}$, $W_i^K \in \mathbb{R}^{d_{model}\times d_k}$, $W_i^V \in \mathbb{R}^{d_{model}\times d_v}$, and $W^O \in \mathbb{R}^{hd_v\times d_{model}}$. $h=8$, $d_k=d_v=\frac{d_{model}}{h}=64$

$$FFN(x)=\max\left(0,xW_1+b_1\right)W_2+b_2 \tag{3}$$

$$\underset{(pos,2i)}{PE}=\sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{4}$$

$$\underset{(pos,2i+1)}{PE}=\cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{5}$$

Transformer(3): Attention is all you need

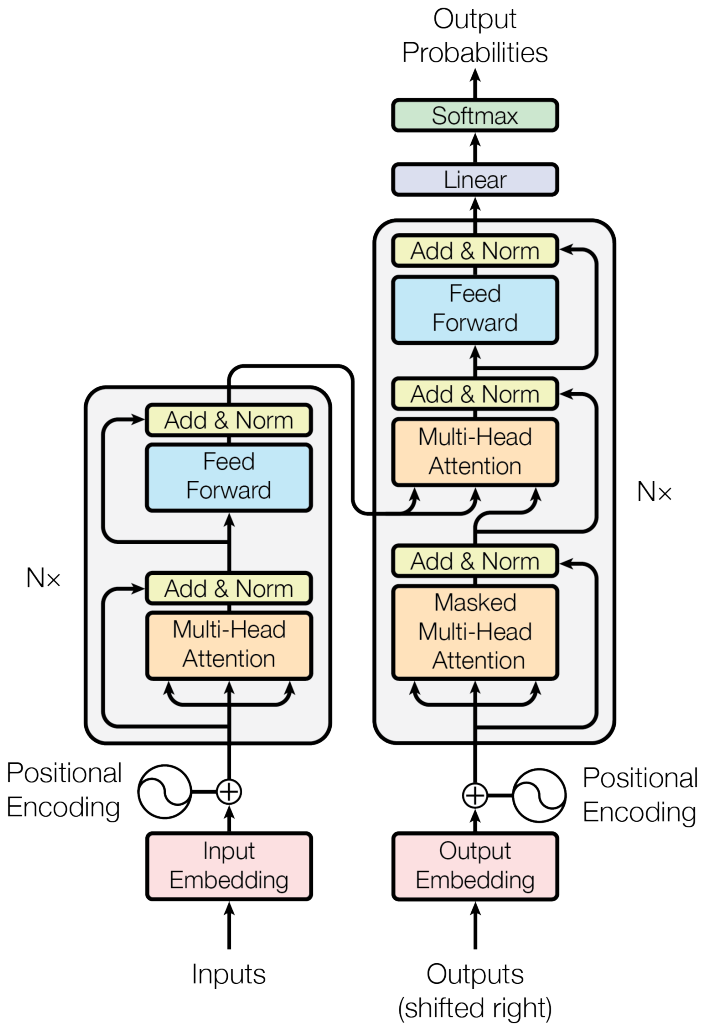


Figure 4: From Vaswani et al. (2017) Fig. 1

BERT, GPT, ELMo 事前訓練の違い

BERT　トランスフォーマー，マスク化言語モデル，次文予測課題

GPT　順方向トランスフォーマー

ELMo　双方向 RNN による中間層の連結

BERT の入力表現

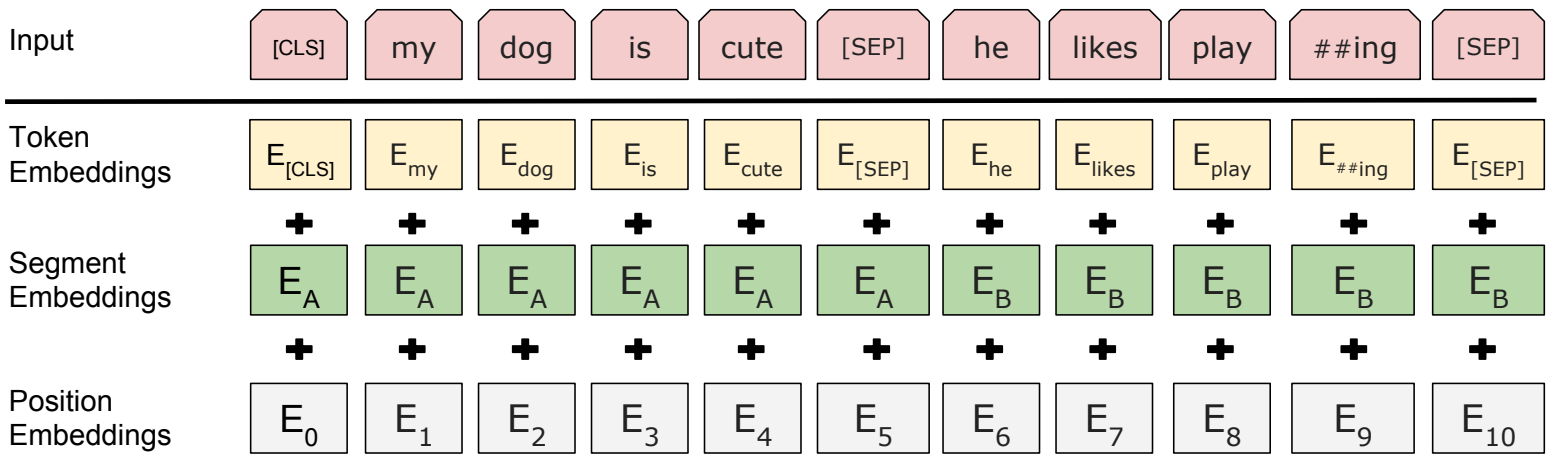


Figure 5: **埋め込みトークンの総和，位置符号器，分離埋め込みの 3 者** From Devlin et al. (2018) Fig. 2

BERT の事前訓練: マスク化言語モデル

全入力系列のうち 15% をランダムに [MASK] トークンで置き換える

- 入力はオリジナル系列を [MASK] トークンで置き換えた系列
- ラベル: オリジナル系列の [MASK] 部分にの正しいラベルを予測

80%: オリジナル入力系列を [MASK] で置換

10%: [MASK] の位置の単語をランダムな無関連語で置き換える

10%: オリジナル系列

BERT の事前訓練: 次文予測課題

言語モデルの欠点を補完する目的，次の文を予測
[SEP] トークンで区切られた 2 文入力

入力: the man went to the store [SEP] he bought a gallon of milk.

ラベル: IsNext

入力: the man went to the store [SEP] penguins are flightless birds.

ラベル: NotNext

BERT: ファインチューニング

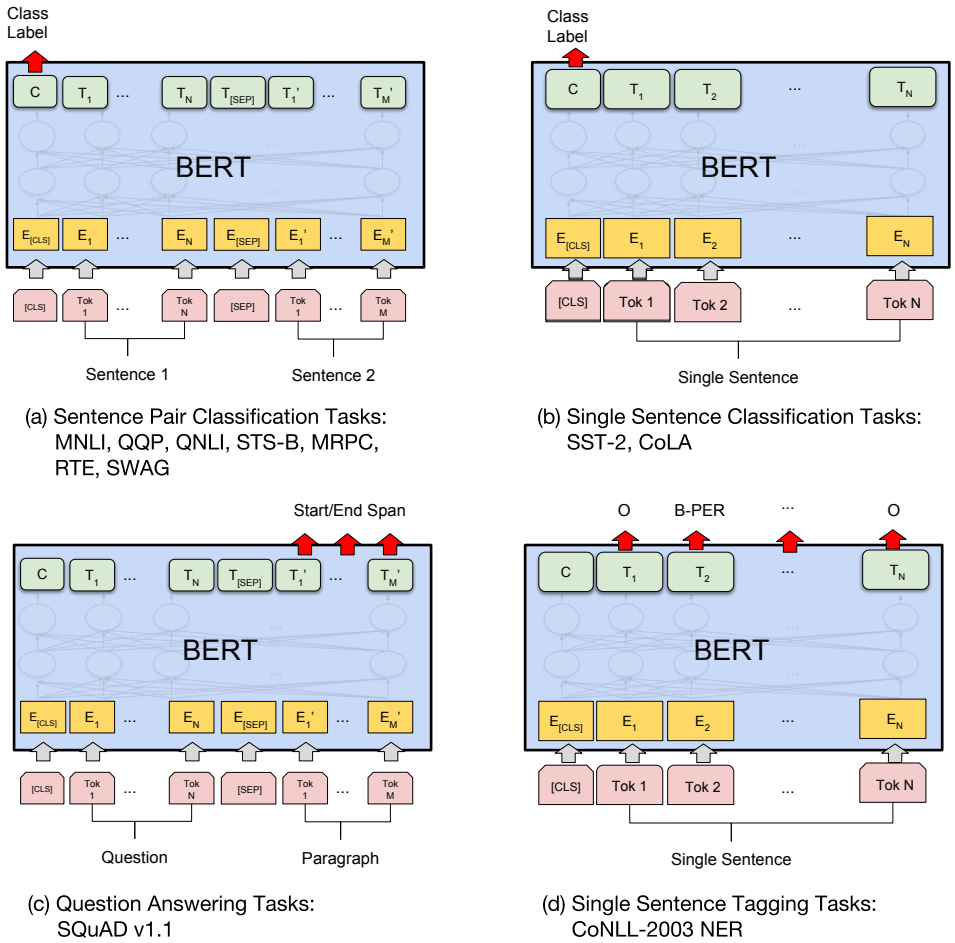


Figure 6: (a), (b) **は文レベル課題，** (c),(d) **はトークンレベル課題**, E: **入力埋め込み表現**, T_i: **トークン** i **の文脈表象。**[CLS]: **分類出力記号**, [SEP]:**文分離記号**. From Devlin et al. (2018) Fig.3

Seq2seq model

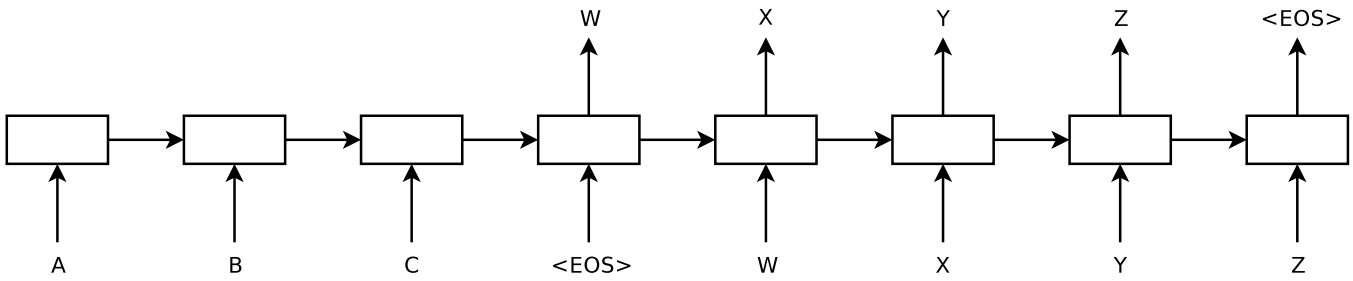


Figure 7: From Sutskever, Vinyals, & Le (2014) Fig. 1 **翻訳モデル** "seq2seq" **の概念図**

"<eos>" **は文末を表す。中央の** "<eos>" **の前がソース言語であり，中央の** "<eos>" **の後はターゲット言語の言語モデルである** SRN **の中間層への入力として用いる。注意すべきは，ソース言語の文終了時の中間層状態のみをターゲット言語の最初の中間層の入力に用いることであり，それ以外の時刻ではソース言語とターゲット言語は関係がない。逆に言えば最終時刻の中間層状態がソース文の情報全てを含んでいるとみなしうる。この点を改善することを目指すことが** 2014 **年以降盛んに行われてきた。顕著な例が後述する 双方向** **RNN**, **LSTM 採用したり，注意 機構を導入することであった。**

Seq2seq (2)

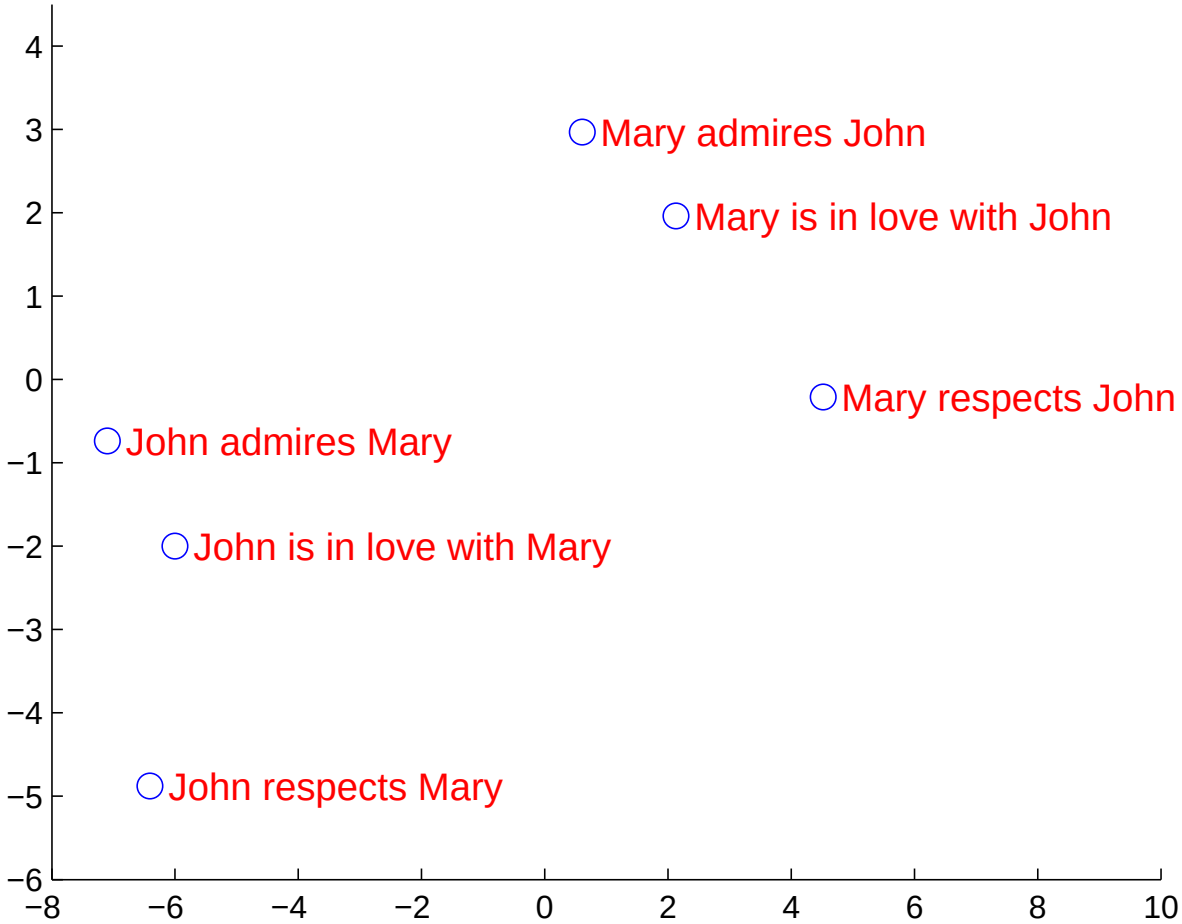


Figure 8: From Sutskever et al. (2014) Fig. 2

Seq2seq (3)

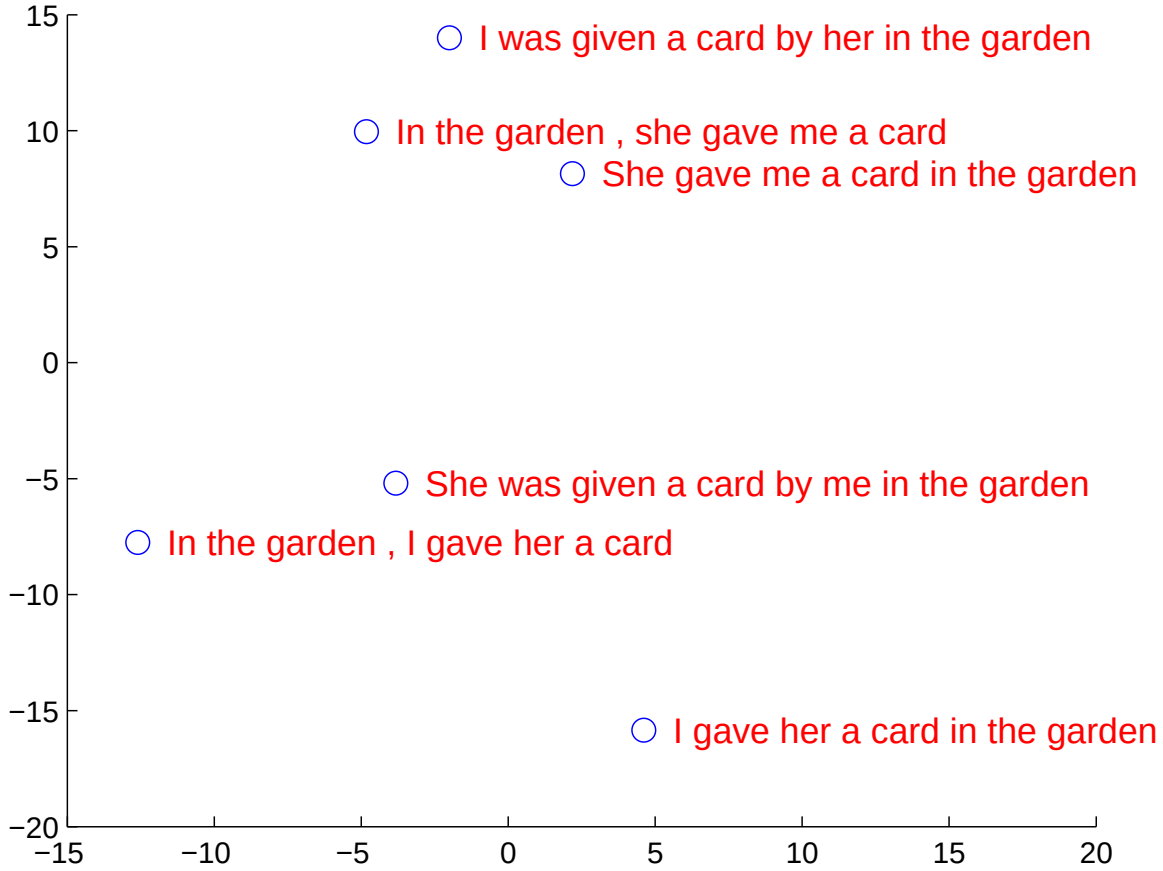


Figure 9: From Sutskever et al. (2014) Fig. 2

自然言語系の注意

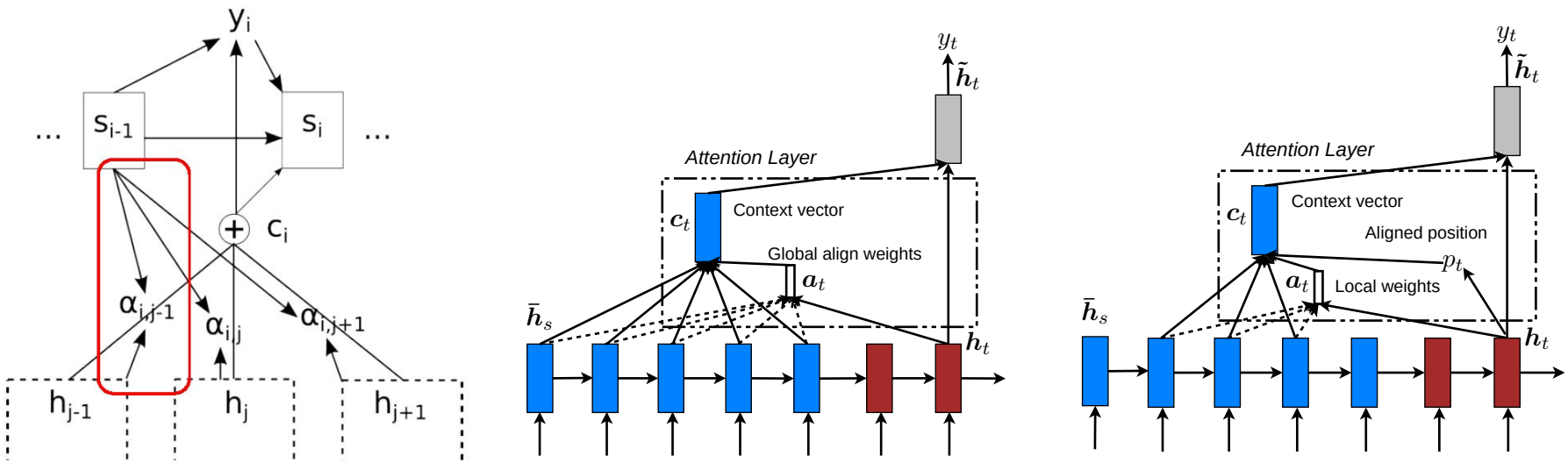


Figure 10: 左:Bahdanau et al. (2015), 中:Luong et al. (2015) Fig. 2, 右:Luong et al. (2015) Fig. 3)

関連分野

- **生理学** Physiology
- **賦活研究** Brain imaging
- **心理学** Psychology
- **神経心理学** Neuropsychology
- **計算モデル** Computational modeling

Dicotomy

- ボトムアップ と トップダウン
- 何 と 何処 (腹側 背側)
- 特徴，対象，場所へ向けられるの注意
- 外発的，内発的 注意

関連脳領域

- FEF **前頭眼野** (Monosov & Thompson, 2009)
- Lateral Intraparietal area (LIP) **側頭頭頂領域** (Wardak, Olivier, & Duhamel, 2004)
- Superior Colliculus(SC) **上丘** (Krauzlis, Lovejoy, & Zénon, 2013)
- PFC **前頭皮質** (Miller & Cohen, 2001)

認知心理学分野

- **フィルタリング** (Broadbent, 1958), **減衰説** (Treisman, 1969)
- **特徴統合理論** (Treisman, 1988; Treisman & Gelade, 1980)
- Guided Search 2.0 (Wolfe, 1994)
- **目標／妨害刺激類似性**: (Duncan & Humphreys, 1989, 1992)
- **サーチライト (スポットライト) 仮説** (Crick, 1984), **ズームレンズ** Eriksen & St.James (1986)
- **勝者占有回路** (Koch & Ullman, 1985) = softmax

計算モデル (Implementation)

- Milanese, Wechsler, Gill, Bost, & Pun (1994)
- Itti, Koch, & Niebur (1998)
- Borji & Itti (2013) SOTA

総説論文

- Itti & Koch (2001)
- Knudsen (2007)
- Petersen & Posner (2012)
- Kimura, Yonetani, & Hirayama (2013)
- Itti & Borji (2015) Oxford Handbook of attention

深層学習系

- **自動翻訳** (Bahdanau et al., 2015; Luong et al., 2015)
- **画像脚注付け** (Vinyals, Toshev, Bengio, & Erhan, 2015)
- **注意** (Wang & Shen, 2018)

ニューラル画像脚注付け

2014 年に提案されたニューラル画像脚注付けのモデル

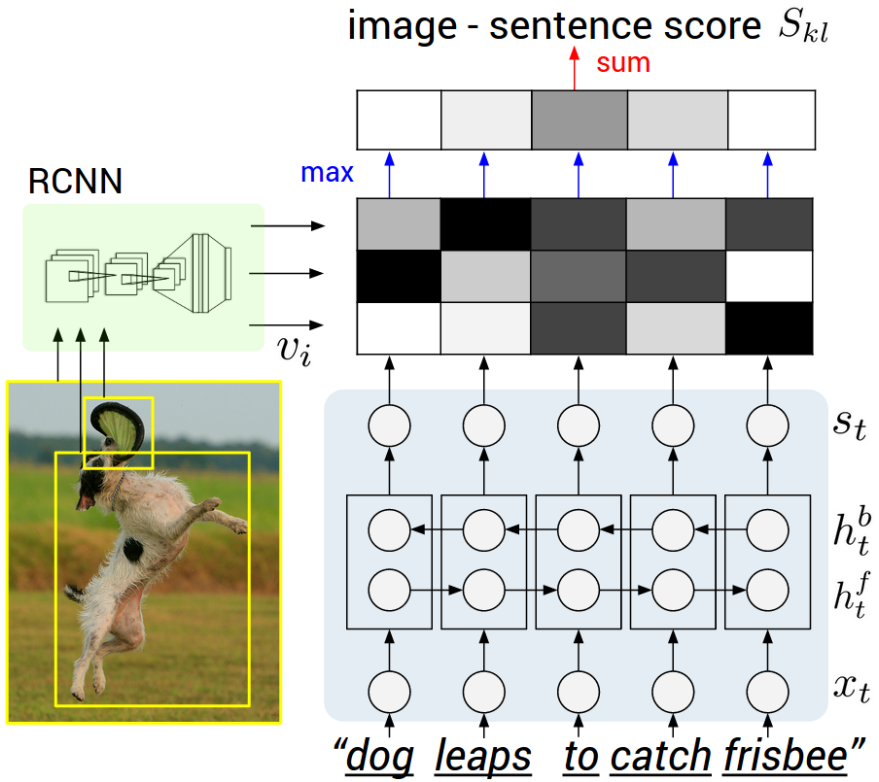


Figure 11: From Karpathy & Fei-Fei (2015) Fig. 3

ニューラル画像脚注付け (2)

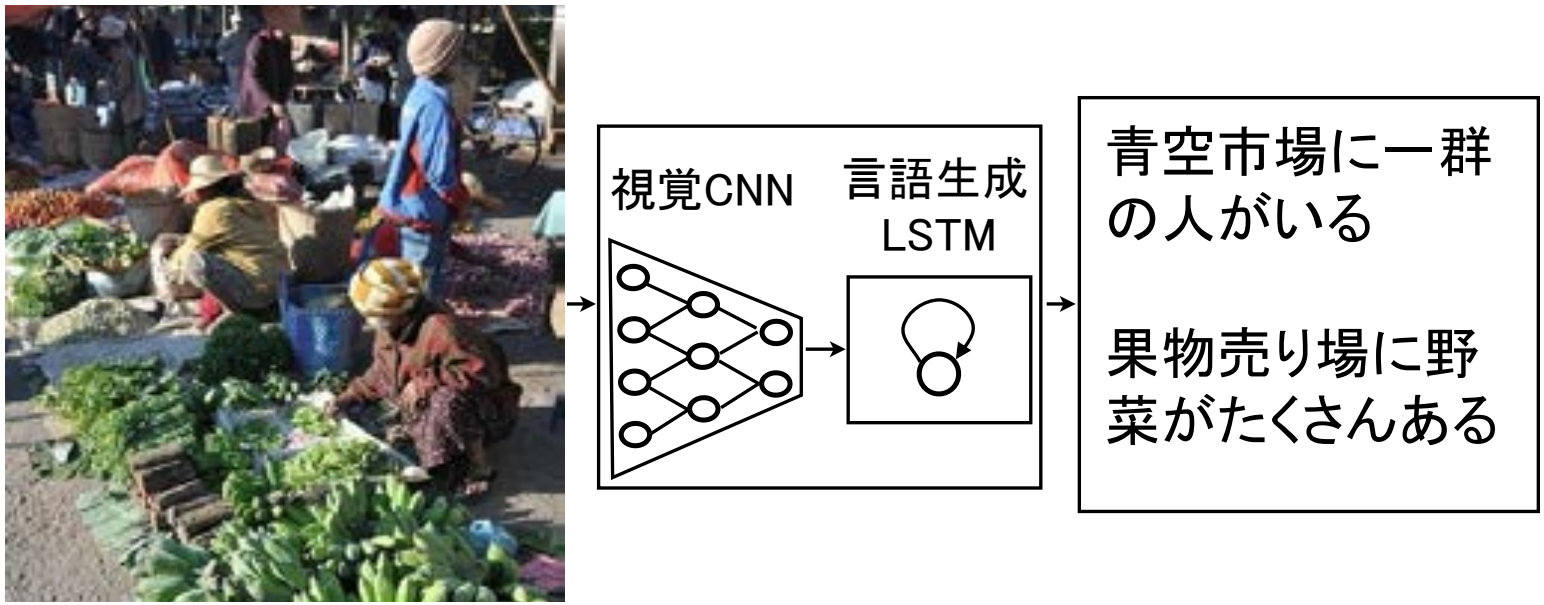


Figure 12: From Vinyals et al. (2015) Fig. 1

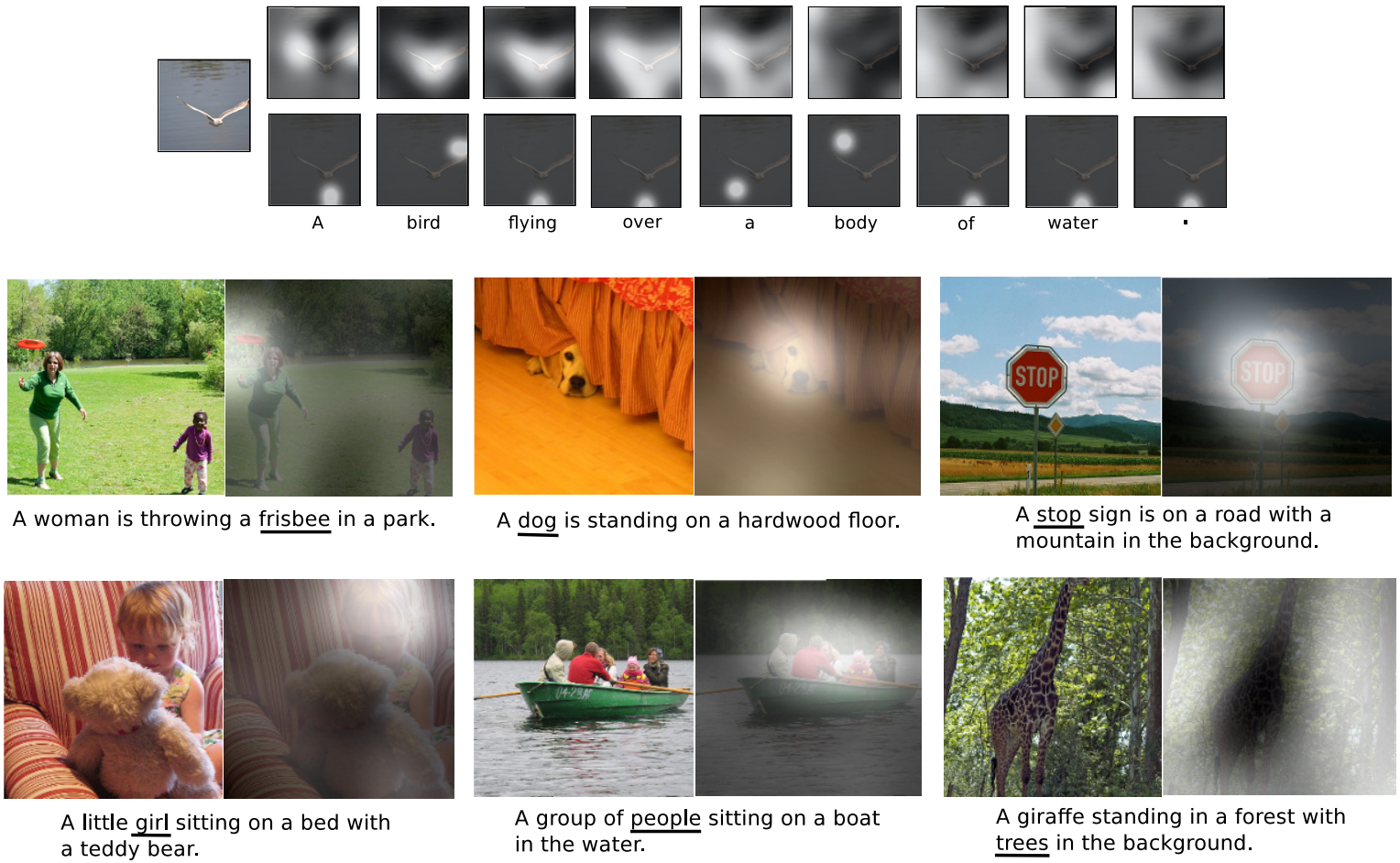


Figure 13: From Xu et al. (2015) Fig. 2

各画像対は右が入力画像であり，左はその入力画像の脚注付けである単語を出力している際にどこに注意しているのかを白色で表している。

温故知新

1. **脳梁切断患者による分離脳** (Sperry, 1961)
2. **半側空間無視** (Heilman & Valenstein, 1979)
3. **頭頂葉損傷患者の注意のディスエンゲージメント** (Posner, 1980)
4. **両耳分離聴実験, カクテルパーティ効果** Broadbent (1958); Treisman (1964)
5. **特徴統合理論** (Treisman, 1988; Treisman & Gelade, 1980)
6. **計算論的モデル サーチライト (スポットライト) 仮説** (Crick, 1984)
7. **モデルとデータセット公開, 競技会** (Itti & Borji, 2014; Itti & Koch, 2001)
8. DeepGazeII (Kummerer et al., 2017)

分離脳 Split brain

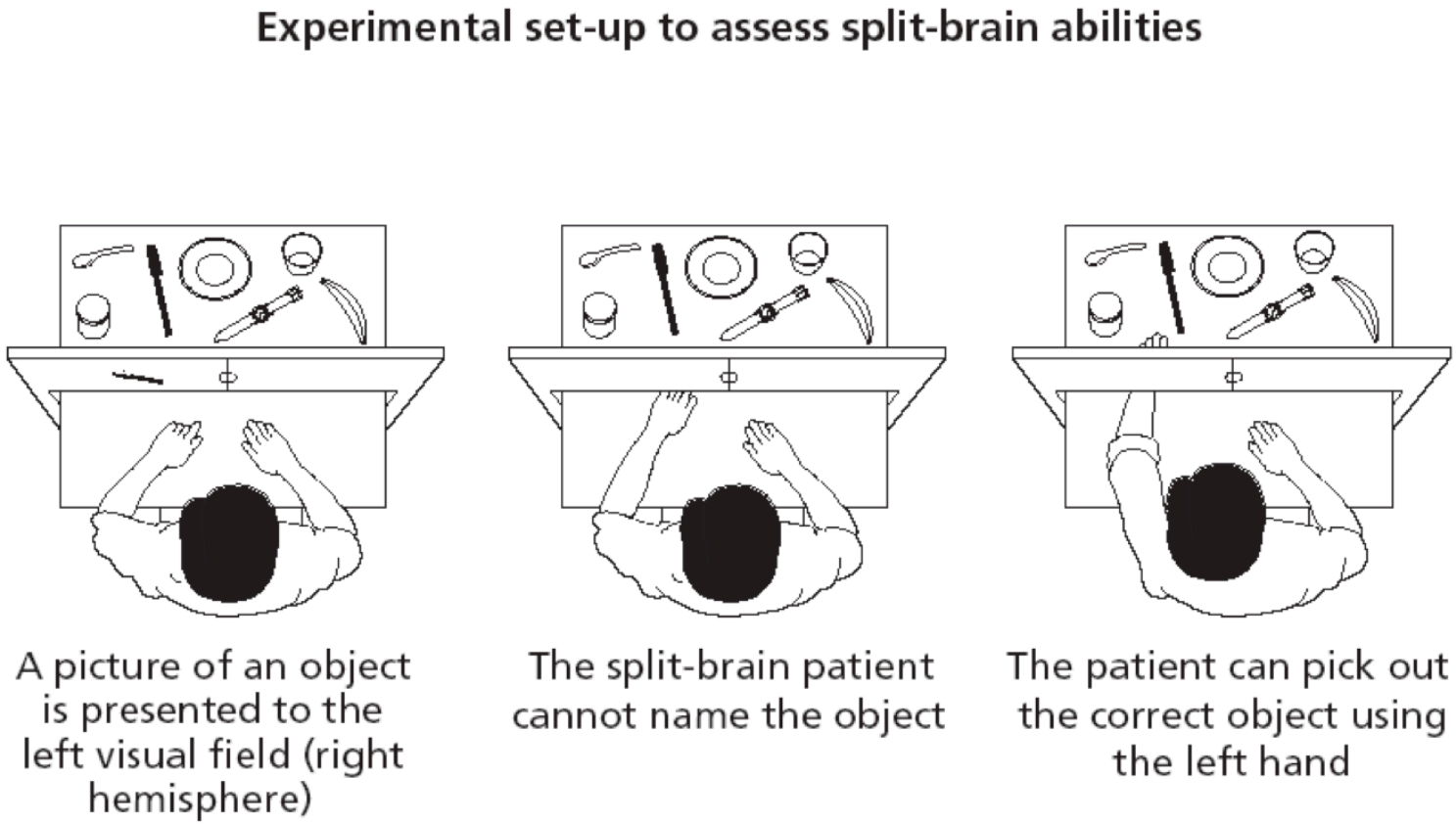


Figure 14: From Sperry (1968) Fig. 5

半側空間無視

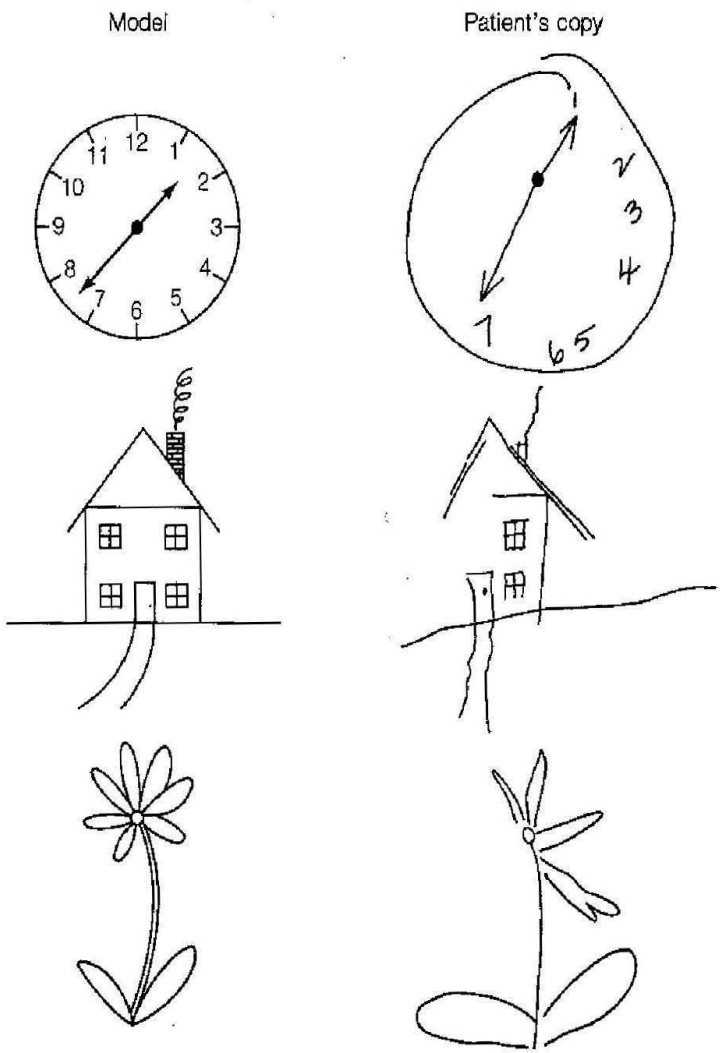


Figure 15: From Bloom & Lazerson (1988) Fig. 17-6

ポズナーとコーヘン

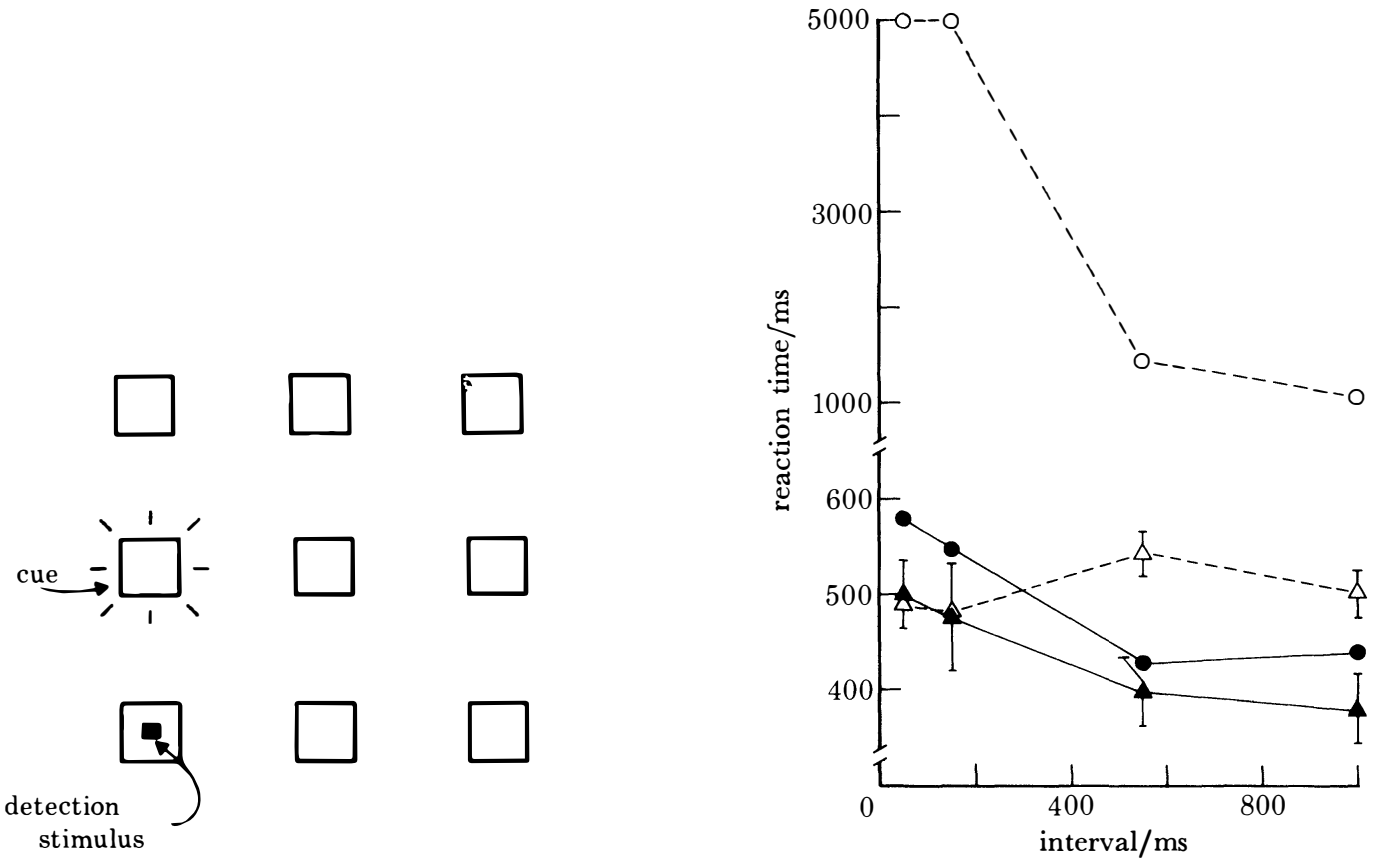


Figure 16: From ? Fig. 1, Fig.6: 右頭頂葉障害を呈した患者 (R.S.) の結果。円:ターゲットが左視野提示，三角:ターゲット右視野提示。白点線:非有効手がかり，黒実線:有効手がかり。横軸は ISI。縦軸は反応時間中央値

特徴統合理論 (FIT)

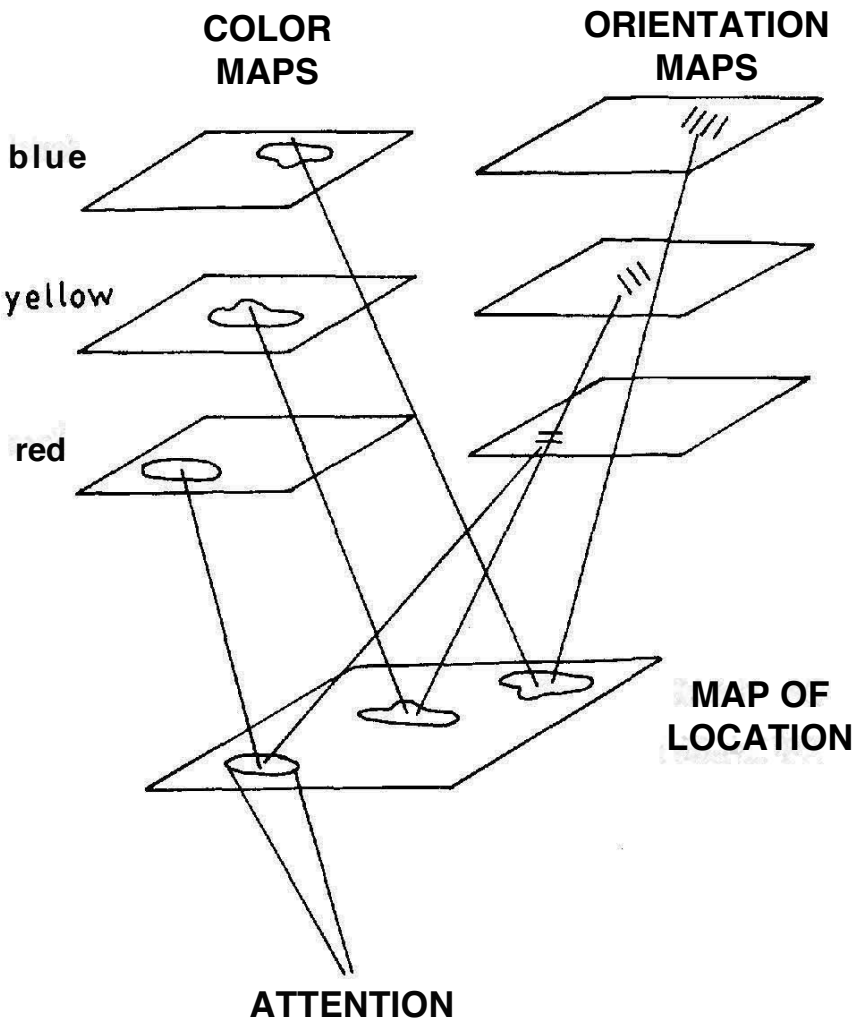


Figure 17: From Treisman & Souther (1985) Fig. 9

探索非对称性 search asymmetry

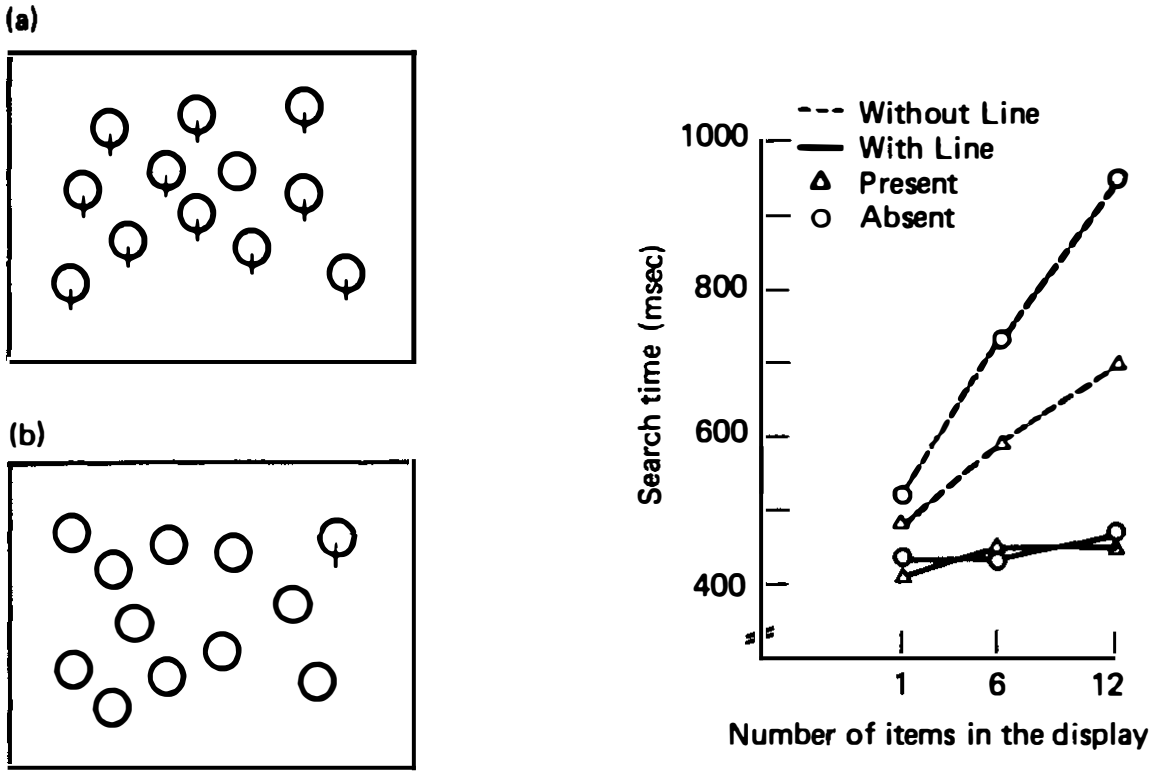


FIG. 3. Examples of displays and mean search times for a target circle with and without an intersecting line.

Figure 18: From Treisman (1988) Fig. 3

スポットライトメタファー

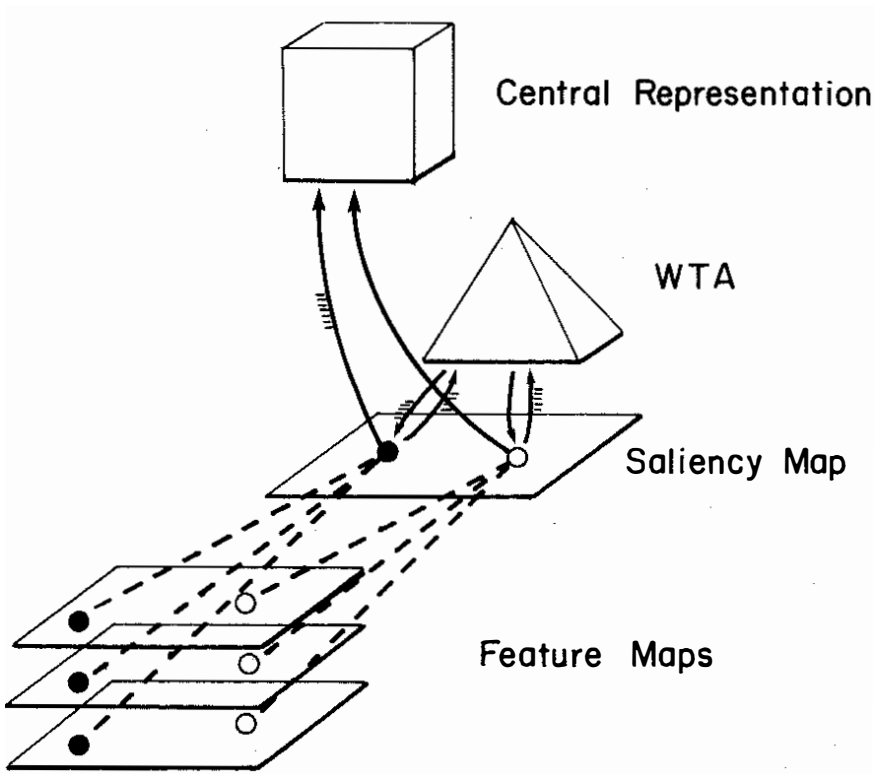


Figure 19: From Koch & Ullman (1985) Fig. 5

ガイド付き探索モデル Guided Search 2.0

最初にトップダウン注意を明示的に示した ガイド付き探索モデル (Wolfe, 1994)

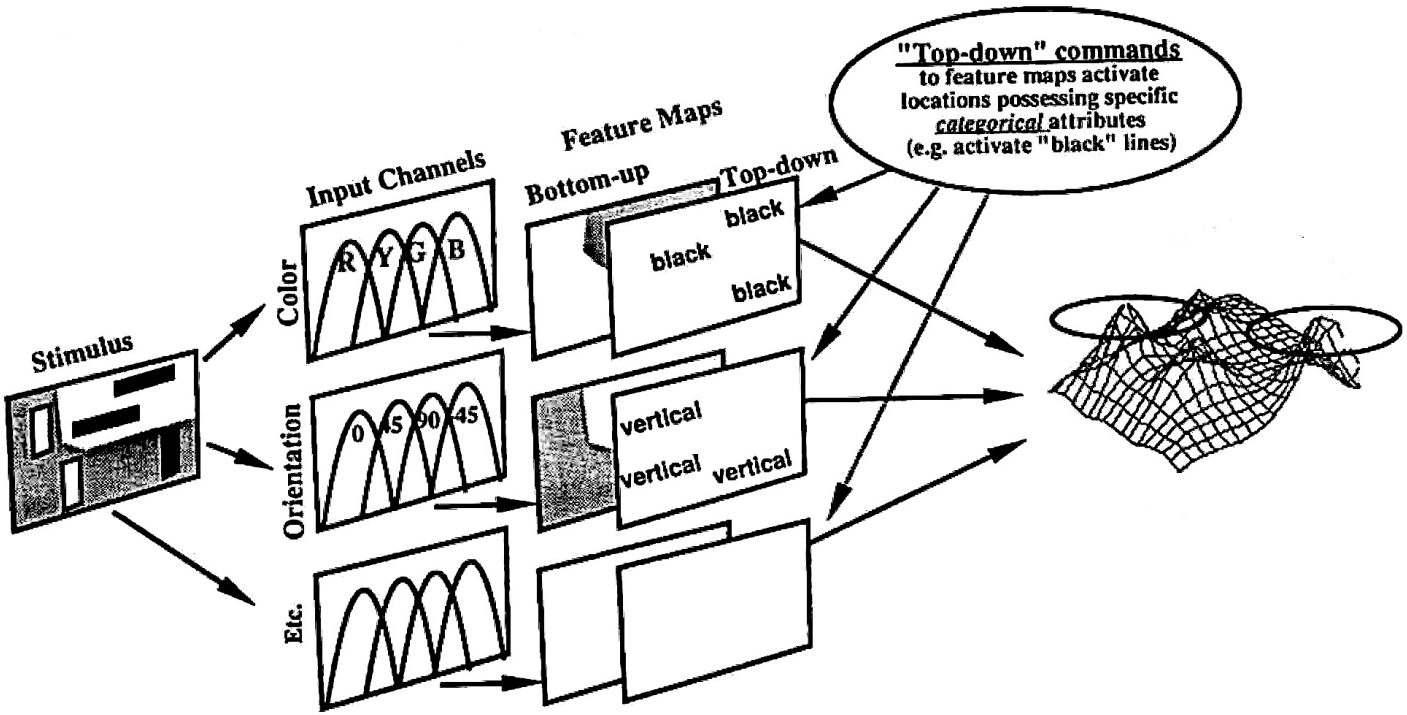


Figure 20: From Wolfe (1994) Fig. 2

Itti & Borji (2015); Itti & Koch (2001) の計算モデル

Itti & Borji (2015) の総説論文からそれまでのモデルの概説図

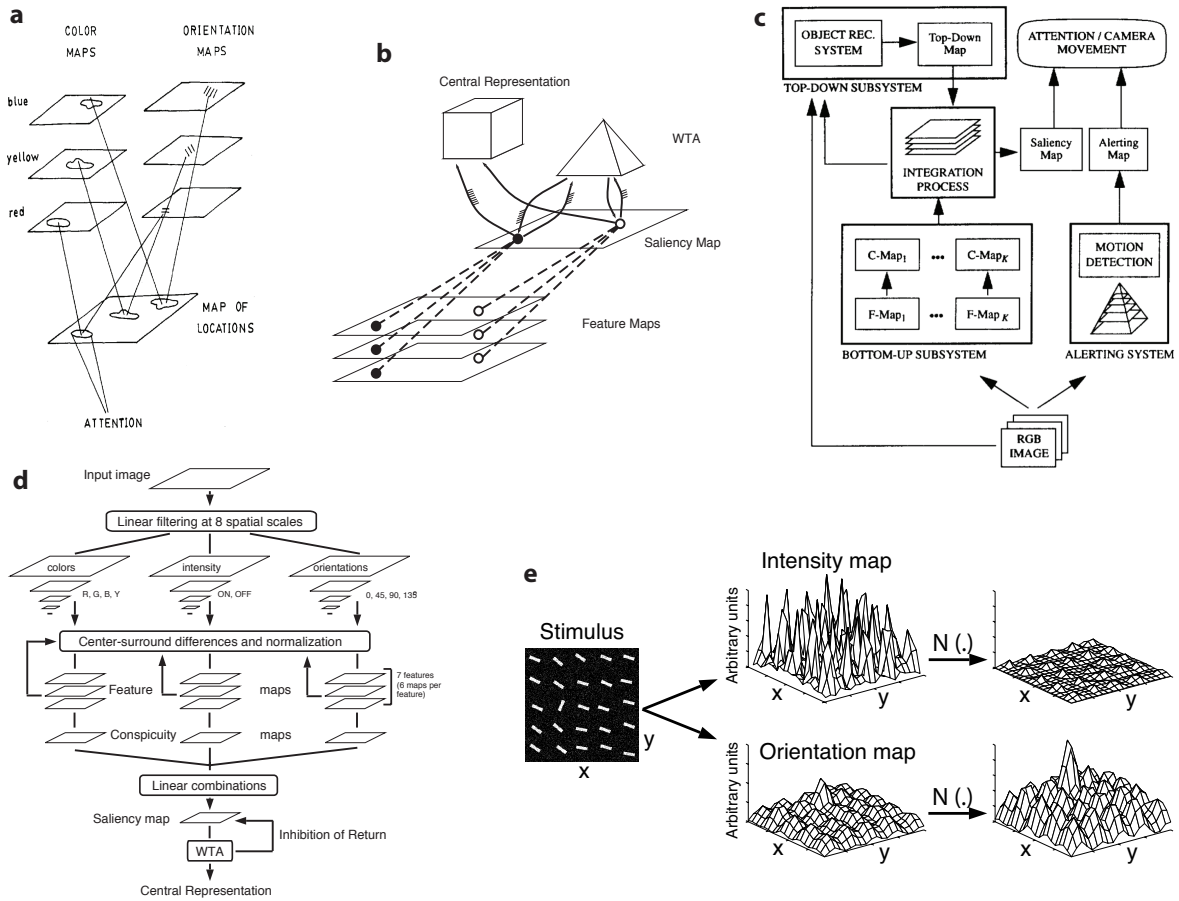
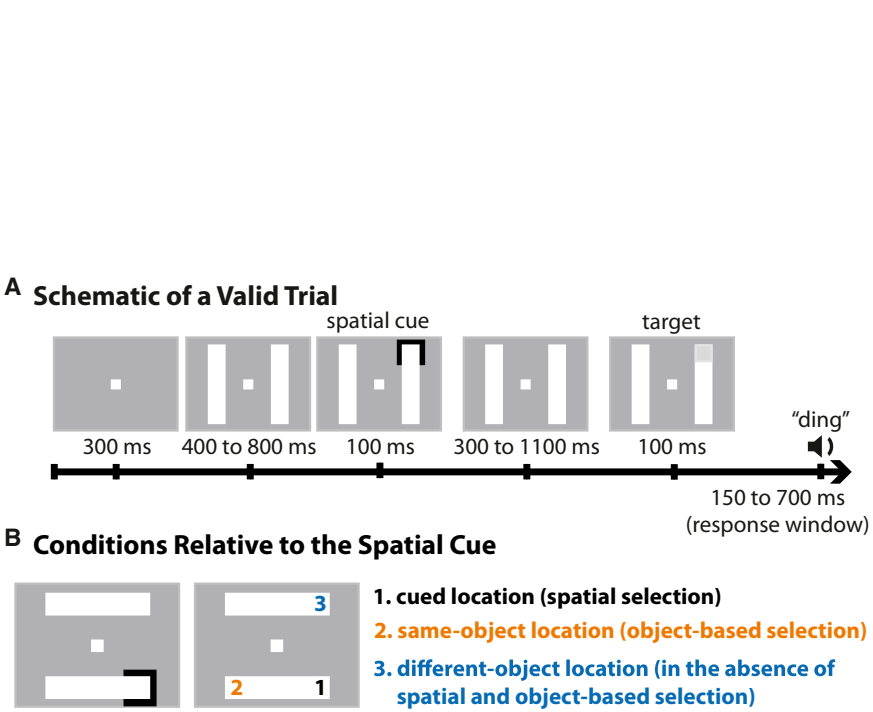


Figure 21: From Itti & Borji (2015) Fig. 2

リズム現象



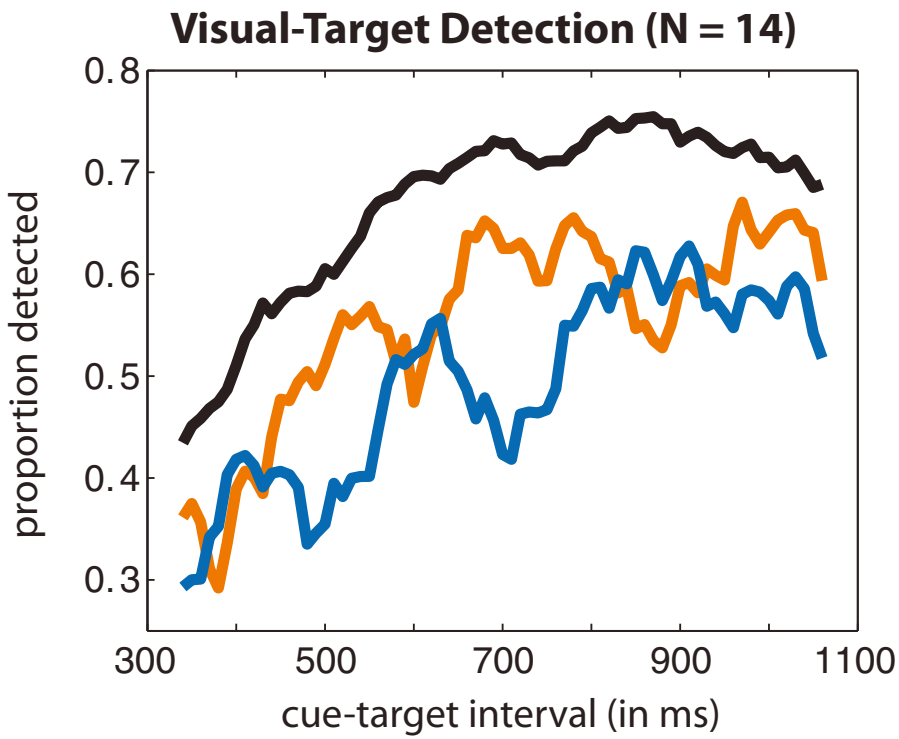


Figure 22: From Fiebelkorn, Saalmann, & Kastner (2013) Fig. 1 and Fig. 2a

リズム現象 (2)

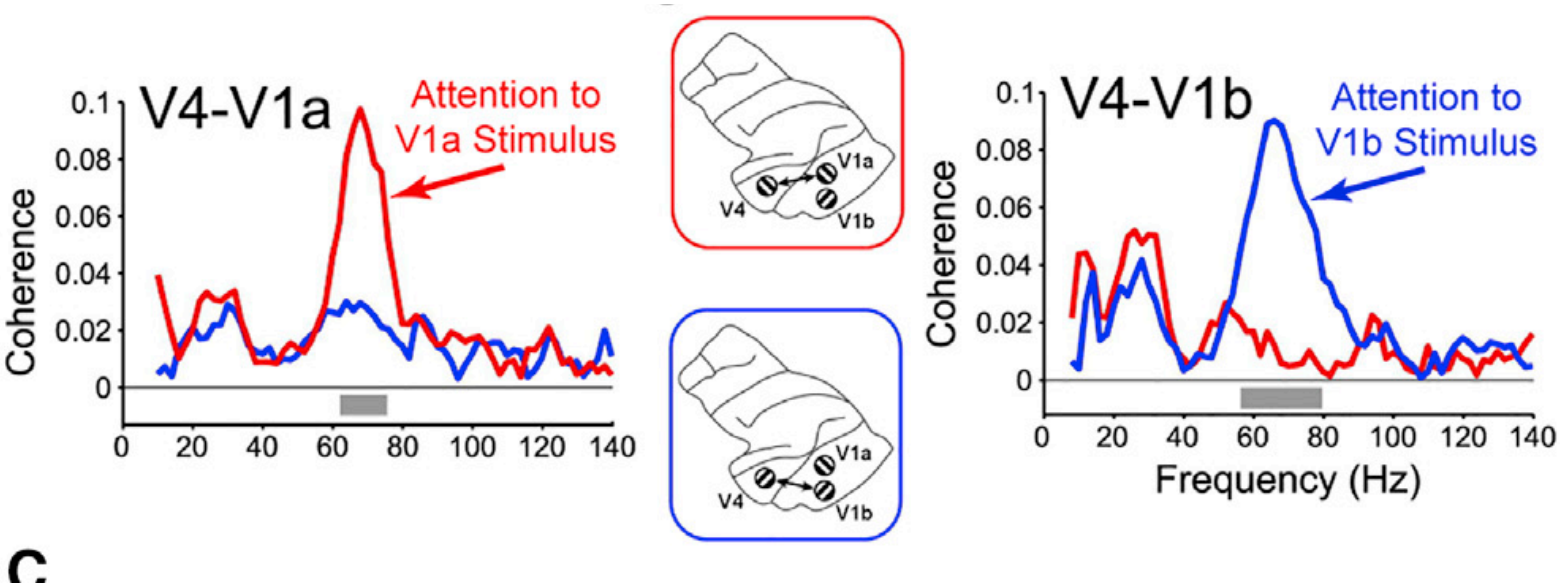


Figure 23: From Buschman & Kastner (2015) Fig. 3b

リズム現象 (3)

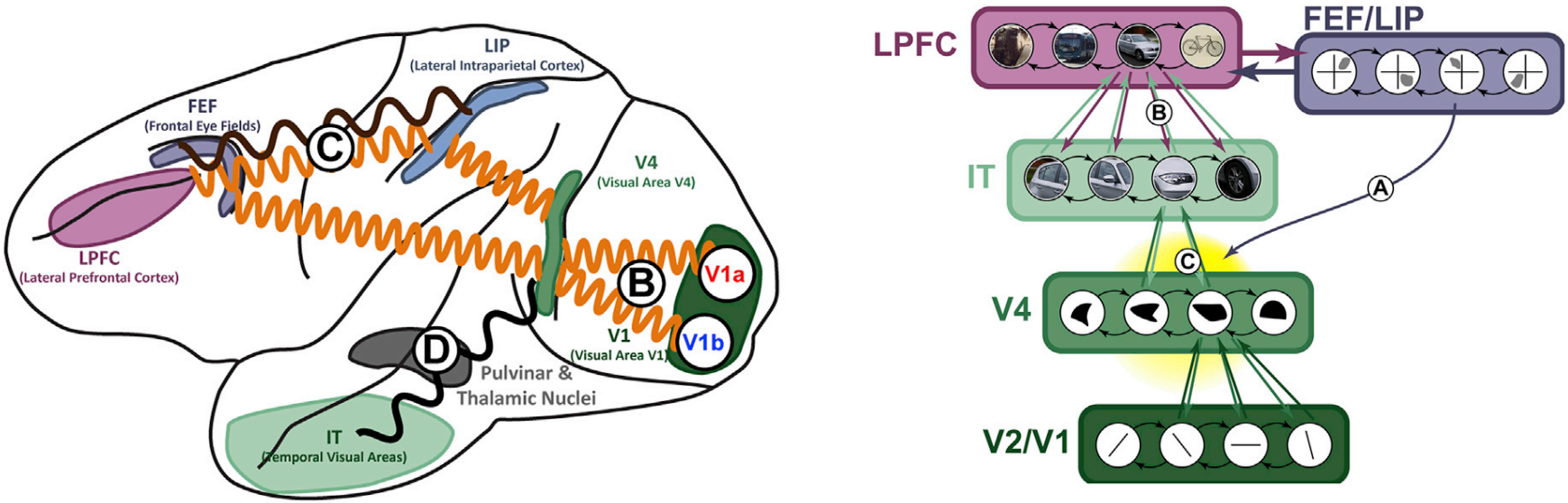


Figure 24: From Buschman & Kastner (2015) Fig. 3a, Fig. 6

データセット

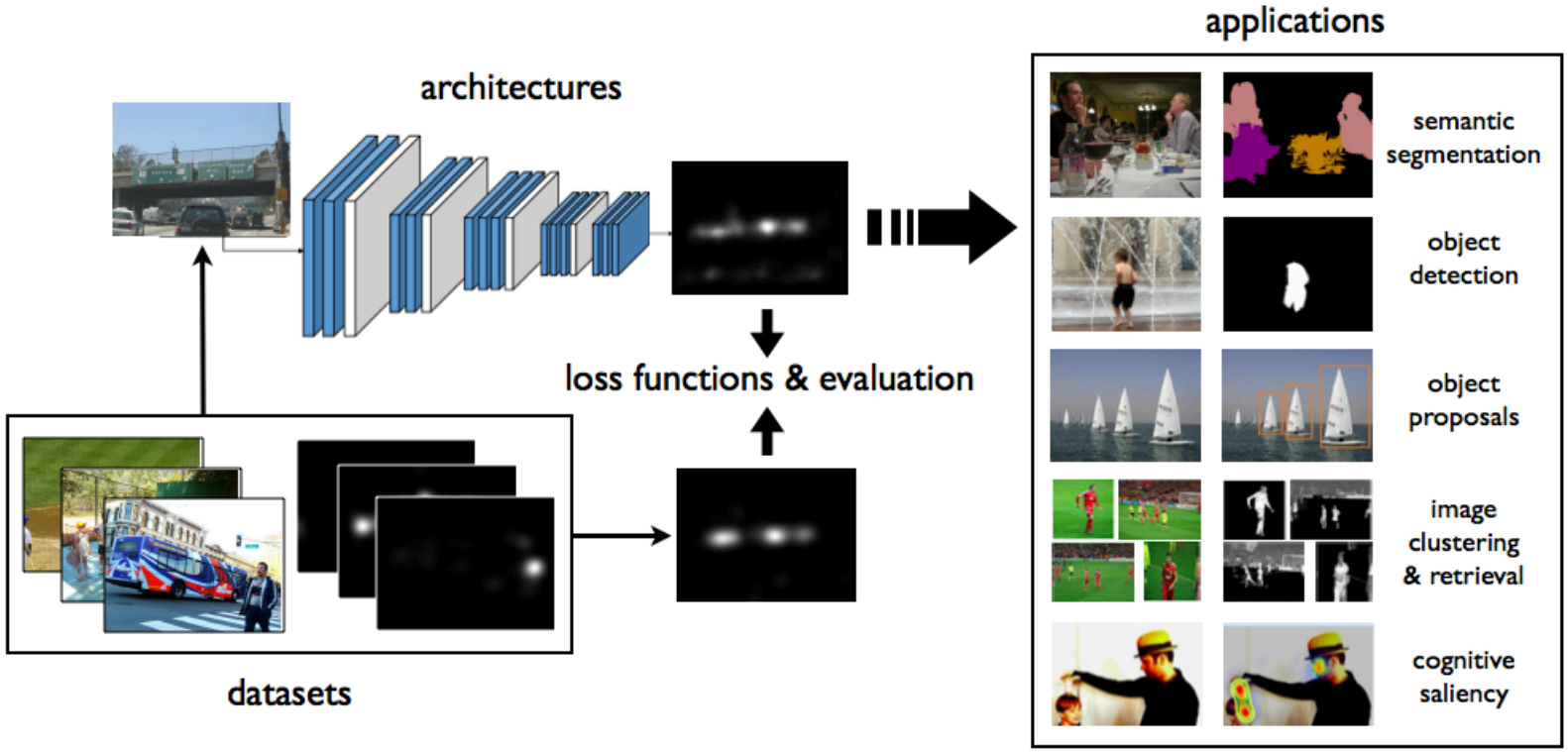


Figure 25: From ECCV2016 tutorial

- MIT300 自然画像 300 枚，被験者 39 名の眼球運動データ
- cat2000 自然画像 2000 枚，被験者 24 名分眼球運動データ
- MIT1003 自然画像 1003 枚

大連工科大学-オムロン データセット

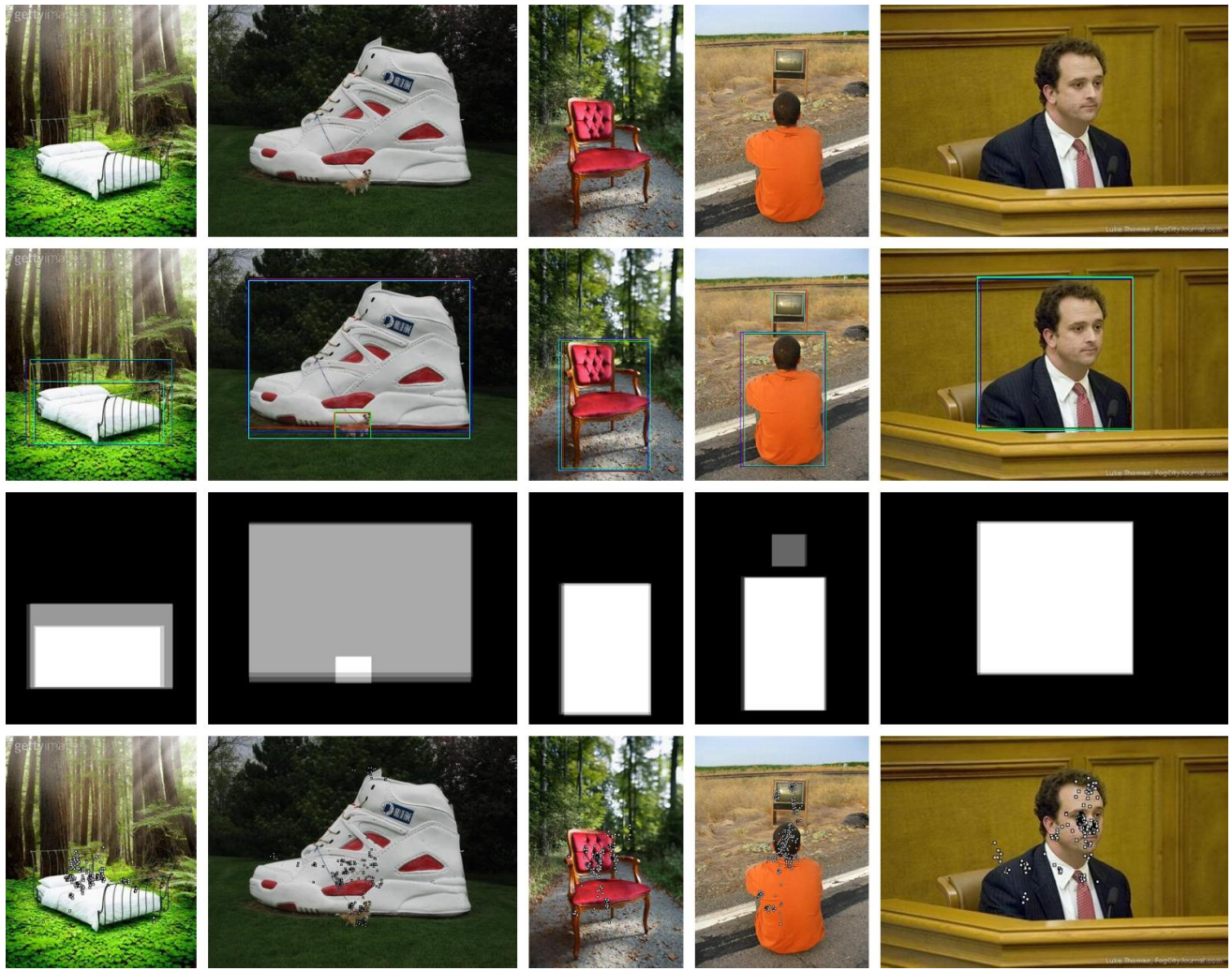
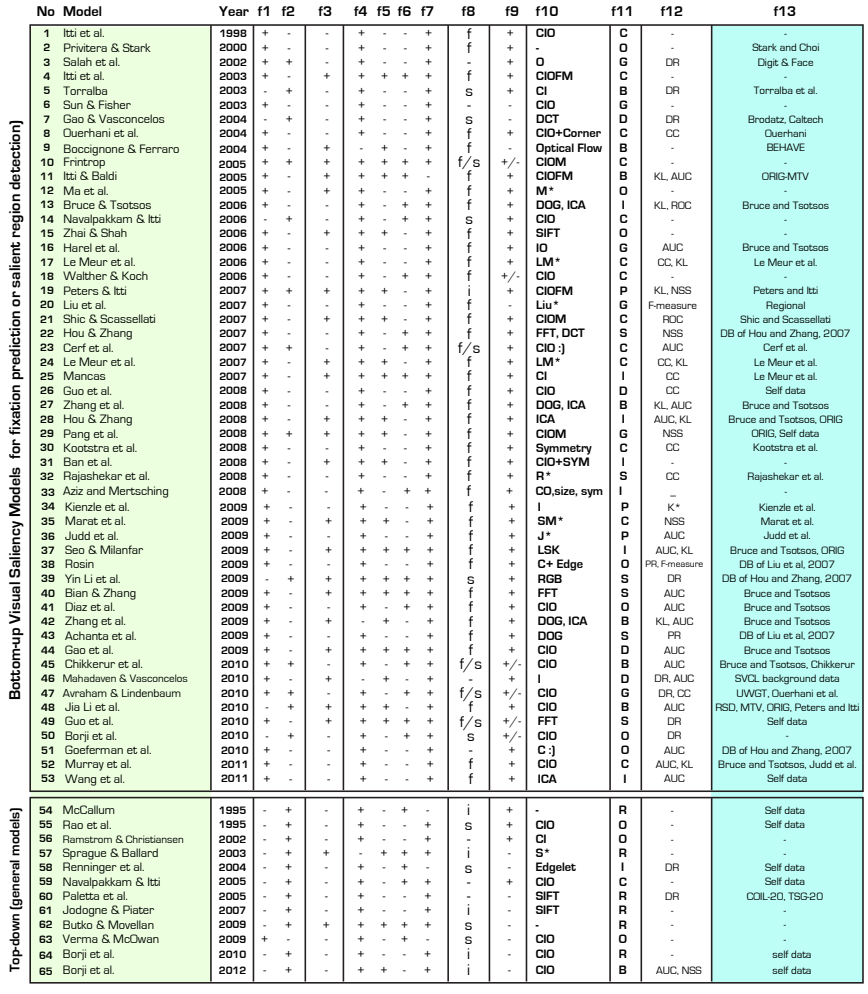


Figure 26: From DUT-OMRON Image dataset



	No	Model	Year	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13
Bottom-up Visual Saliency Models for fixation prediction or salient region detection]	1	Itti et al.	1998	+	-	-	+	-	-	+	f	+	CIO	C	-	-
	2	Privitera & Stark	2000	+	-	-	+	-	-	+	f	+	-	O	-	Stark and Choi
	3	Salah et al.	2002	+	+	-	+	-	-	+	-	+	O	G	DR	Digit & Face
	4	Itti et al.	2003	+	-	+	+	+	+	+	f	+	CIOFM	C	-	-
	5	Torralba	2003	-	+	-	+	-	-	+	s	+	CI	B	DR	Torralba et al.
	6	Sun & Fisher	2003	+	-	-	+	-	-	+	-	-	CIO	G	-	-
	7	Gao & Vasconcelos	2004	-	+	-	+	-	-	+	s	-	DCT	D	DR	Brodatz, Caltech
	8	Ouerhani et al.	2004	+	-	-	+	-	-	+	f	+	CIO+Corner	C	CC	Ouerhani
	9	Boccignone & Ferraro	2004	+	-	+	-	+	-	+	f	-	Optical Flow	B	-	BEHAVE
	10	Frintrop	2005	+	+	+	+	+	+	+	f/s	+/-	CIOM	C	-	-
	11	Itti & Baldi	2005	+	-	+	+	+	+	-	f	+	CIOFM	B	KL, AUC	ORIG-MTV
	12	Ma et al.	2005	+	-	+	+	-	-	+	f	+	M*	O	-	-
	13	Bruce & Tsotsos	2006	+	-	-	+	-	+	+	f	+	DOG, ICA	I	KL, ROC	Bruce and Tsotsos
	14	Navalpakkam & Itti	2006	-	+	-	+	-	+	+	s	+	CIO	C	-	-
	15	Zhai & Shah	2006	+	-	+	+	+	-	+	f	+	SIFT	O	-	-
	16	Harel et al.	2006	+	-	-	+	-	-	+	f	+	IO	G	AUC	Bruce and Tsotsos
	17	Le Meur et al.	2006	+	-	-	+	-	-	+	f	+	LM*	C	CC, KL	Le Meur et al.
	18	Walther & Koch	2006	+	-	-	+	-	+	+	f	+/-	CIO	C	-	-
	19	Peters & Itti	2007	+	+	+	+	+	-	+	i	+	CIOFM	P	KL, NSS	Peters and Itti
	20	Liu et al.	2007	+	-	-	+	-	-	+	f	-	Liu*	G	F-measure	Regional
	21	Shic & Scassellati	2007	+	-	+	+	+	-	+	f	+	CIOM	C	ROC	Shic and Scassellati
	22	Hou & Zhang	2007	+	-	-	+	-	+	+	f	+	FFT, DCT	S	NSS	DB of Hou and Zhang, 2007
	23	Cerf et al.	2007	+	+	-	+	-	+	+	f/s	+	CIO :)	C	AUC	Cerf et al.
	24	Le Meur et al.	2007	+	-	+	+	+	-	+	f	+	LM*	C	CC, KL	Le Meur et al.
	25	Mancas	2007	+	-	+	+	+	+	+	f	+	CI	I	CC	Le Meur et al.
	26	Guo et al.	2008	+	-	-	+	-	-	+	f	+	CIO	D	CC	Self data
	27	Zhang et al.	2008	+	-	-	+	-	+	+	f	+	DOG, ICA	B	KL, AUC	Bruce and Tsotsos
	28	Hou & Zhang	2008	+	-	+	+	+	-	+	f	+	ICA	I	AUC, KL	Bruce and Tsotsos, ORIG
	29	Pang et al.	2008	+	+	+	+	+	-	+	f	+	CIOM	G	NSS	ORIG, Self data
	30	Kootstra et al.	2008	+	-	-	+	-	-	+	f	+	Symmetry	C	CC	Kootstra et al.
	31	Ban et al.	2008	+	-	+	+	+	-	+	f	+	CIO+SYM	I	-	-
	32	Rajashekar et al.	2008	+	-	-	+	-	-	+	f	+	R*	S	CC	Rajashekar et al.
	33	Aziz and Mertsching	2008	+	-	-	+	-	+	+	f	+	CO,size, sym	I	-	-
	34	Kienzle et al.	2009	+	-	-	+	-	-	+	f	+	I	P	K*	Kienzle et al.
	35	Marat et al.	2009	+	-	+	+	+	-	+	f	+	SM*	C	NSS	Marat et al.
	36	Judd et al.	2009	+	-	-	+	-	-	+	f	+	J*	P	AUC	Judd et al.
	37	Seo & Milanfar	2009	+	-	+	+	+	+	+	f	+	LSK	I	AUC, KL	Bruce and Tsotsos, ORIG
	38	Rosin	2009	+	-	-	+	-	-	+	f	+	C+ Edge	O	PR, F-measure	DB of Liu et al, 2007
	39	Yin Li et al.	2009	-	+	+	+	+	+	+	s	+	RGB	S	DR	DB of Hou and Zhang, 2007
	40	Bian & Zhang	2009	+	-	+	+	+	+	+	f	+	FFT	S	AUC	Bruce and Tsotsos
	41	Diaz et al.	2009	+	-	-	+	-	+	+	f	+	CIO	O	AUC	Bruce and Tsotsos
	42	Zhang et al.	2009	+	-	+	-	+	-	+	f	+	DOG, ICA	B	KL, AUC	Bruce and Tsotsos
	43	Achanta et al.	2009	+	-	-	+	-	-	+	f	+	DOG	S	PR	DB of Liu et al, 2007
	44	Gao et al.	2009	+	-	+	+	+	+	+	f	+	CIO	D	AUC	Bruce and Tsotsos
	45	Chikkerur et al.	2010	+	+	-	+	-	+	+	f/s	+/-	CIO	B	AUC	Bruce and Tsotsos, Chikkerur
	46	Mahadaven & Vasconcelos	2010	+	-	+	-	+	-	+	-	+	I	D	DR, AUC	SVCL background data
	47	Avraham & Lindenbaum	2010	+	+	-	+	-	+	+	f/s	+/-	CIO	G	DR, CC	UWGT, Ouerhani et al.
	48	Jia Li et al.	2010	-	+	+	+	+	-	+	f	+	CIO	B	AUC	RSD, MTV, ORIG, Peters and Itti
	49	Guo et al.	2010	+	-	+	+	+	+	+	f/s	+/-	FFT	S	DR	Self data
	50	Borji et al.	2010	-	+	-	+	-	+	+	s	+/-	CIO	O	DR	-
	51	Goeferman et al.	2010	+	-	-	+	-	-	+	-	+	C :)	O	AUC	DB of Hou and Zhang, 2007
	52	Murray et al.	2011	+	-	-	+	-	-	+	f	+	CIO	C	AUC, KL	Bruce and Tsotsos, Judd et al.
	53	Wang et al.	2011	+	-	-	+	-	-	+	f	+	ICA	I	AUC	Self data
Top-down (general models)	54	McCallum	1995	-	+	-	+	-	+	-	i	+	-	R	-	Self data
	55	Rao et al.	1995	-	+	-	+	-	-	+	s	+	CIO	O	-	Self data
	56	Ramstrom & Christiansen	2002	-	+	-	+	-	-	+	-	+	CI	O	-	-
	57	Sprague & Ballard	2003	-	+	+	-	+	+	+	i	-	S*	R	-	-
	58	Renninger et al.	2004	-	+	-	+	-	+	-	s	-	Edgelet	I	DR	Self data
	59	Navalpakkam & Itti	2005	-	+	-	+	-	+	+	-	+	CIO	C	-	Self data
	60	Paletta et al.	2005	-	+	-	+	-	-	+	-	-	SIFT	R	DR	COIL-20, TSG-20
	61	Jodogne & Piater	2007	-	+	-	+	-	-	+	i	-	SIFT	R	-	-
	62	Butko & Movellan	2009	-	+	+	+	+	+	+	s	-	-	R	-	-
	63	Verma & McOwan	2009	+	-	-	+	-	+	-	s	-	CIO	O	-	-
	64	Borji et al.	2010	-	+	-	+	-	-	+	i	-	CIO	R	-	self data
	65	Borji et al.	2012	-	+	-	+	+	-	+	i	-	CIO	B	AUC, NSS	self data

Figure 3: Survey of bottom-up and top-down computational models, classified according to 13 factors. Factors in order are: Bottom-up (f_1), Top-down (f_2), Spatial (-)/Spatio-temporal (+) (f_3), Static (f_4), Dynamic (f_5), Synthetic (f_6) and Natural (f_7) stimuli, Task-type (f_8), Space-based(+)/Object-based(-) (f_9), Features (f_{10}), Model type (f_{11}), Measures (f_{12}), and Used dataset (f_{13}). In Task type (f_8) column: free-viewing (f); target search (s); interactive (i). In Features (f_{10}) column: CIO: color, intensity and orientation saliency; CIOFM: CIO plus flicker and motion saliency; M* = motion saliency, static saliency, camera motion, object (face) and aural saliency (Speech-music); LM* = contrast sensitivity, perceptual decomposition, visual masking and center-surround interactions; Liu* = center-surround histogram, multi-scale contrast and color spatial-distribution; R* = luminance, contrast, luminance-bandpass, contrast-bandpass; SM* = orientation and motion; J* = CIO, horizontal line, face, people detector, gist, etc; S* = color matching, depth and lines; :) = face. In Model type (f_{11}) column, R means that a model is based RL. In Measures (f_{12}) column: K* = used Wilcoxon-Mann-Whitney test (The probability that a random chosen target patch receives higher saliency than a randomly chosen negative one); DR means that models have used a measure of detection/classification rate to determine how successful was a model. PR stands for Precision-Recall. In dataset (f_{13}) column: Self data means that authors gathered their own data. For a detailed definition of these factors please refer to Borji & Itti (2012 PAMI).

Figure 27: From Itti & Borji (2015) Fig. 3

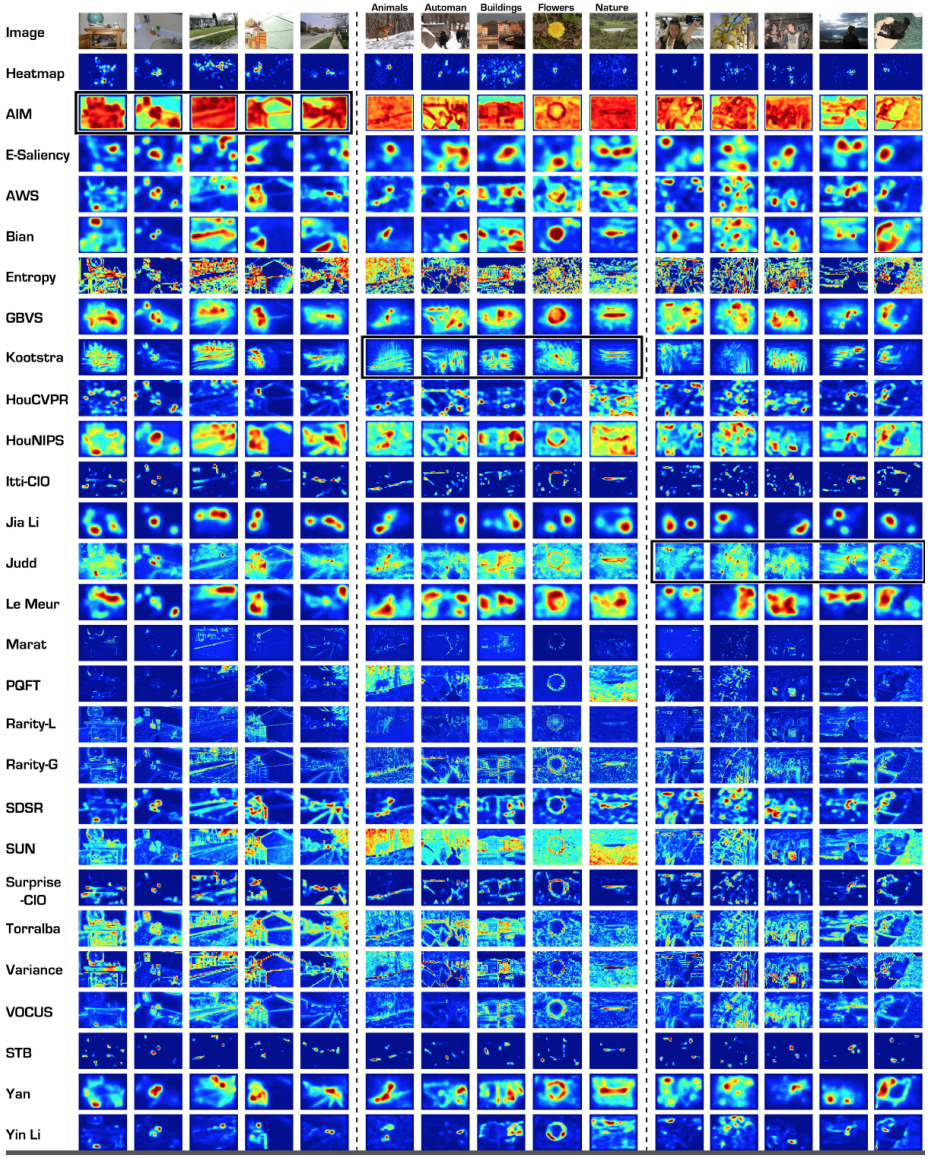


Figure 4: Example images (first row), human eye movement heatmaps (second row), and saliency maps from 26 computational models. The three vertical dashed lines separate the three datasets used (Bruce & Tsotsos, Kootstra & Schomacker, and Judd *et al.*). Black rectangles indicate the model originally associated with a given image dataset. Please see Borji *et al.* (2012 TIP) for additional details.

Figure 28: From Itti & Borji (2015) Fig. 4

DeepGaze II

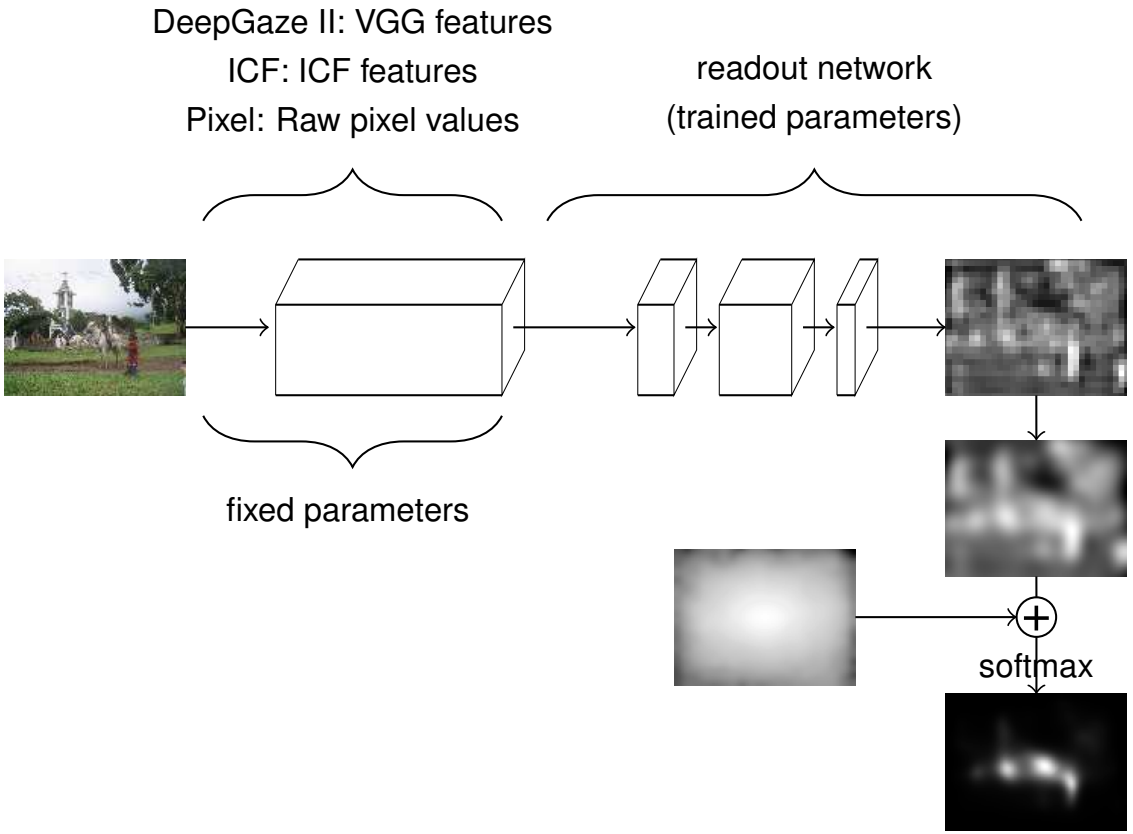


Figure 29: From Kummerer et al. (2017) Fig. 2

DeepGaze II (2)

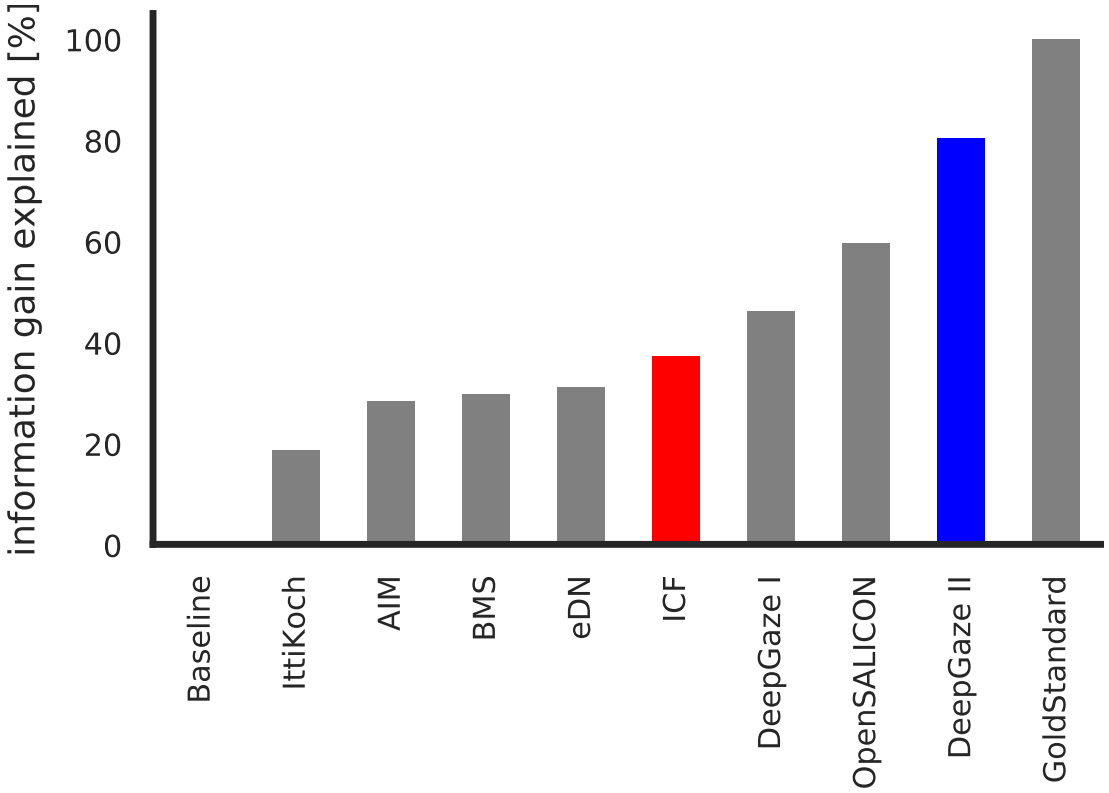


Figure 30: From Kummerer et al. (2017) Fig. 2

DeepGazeII **より成績の良い最右の棒は人間の眼球運動データ**

DeepGaze II (3)

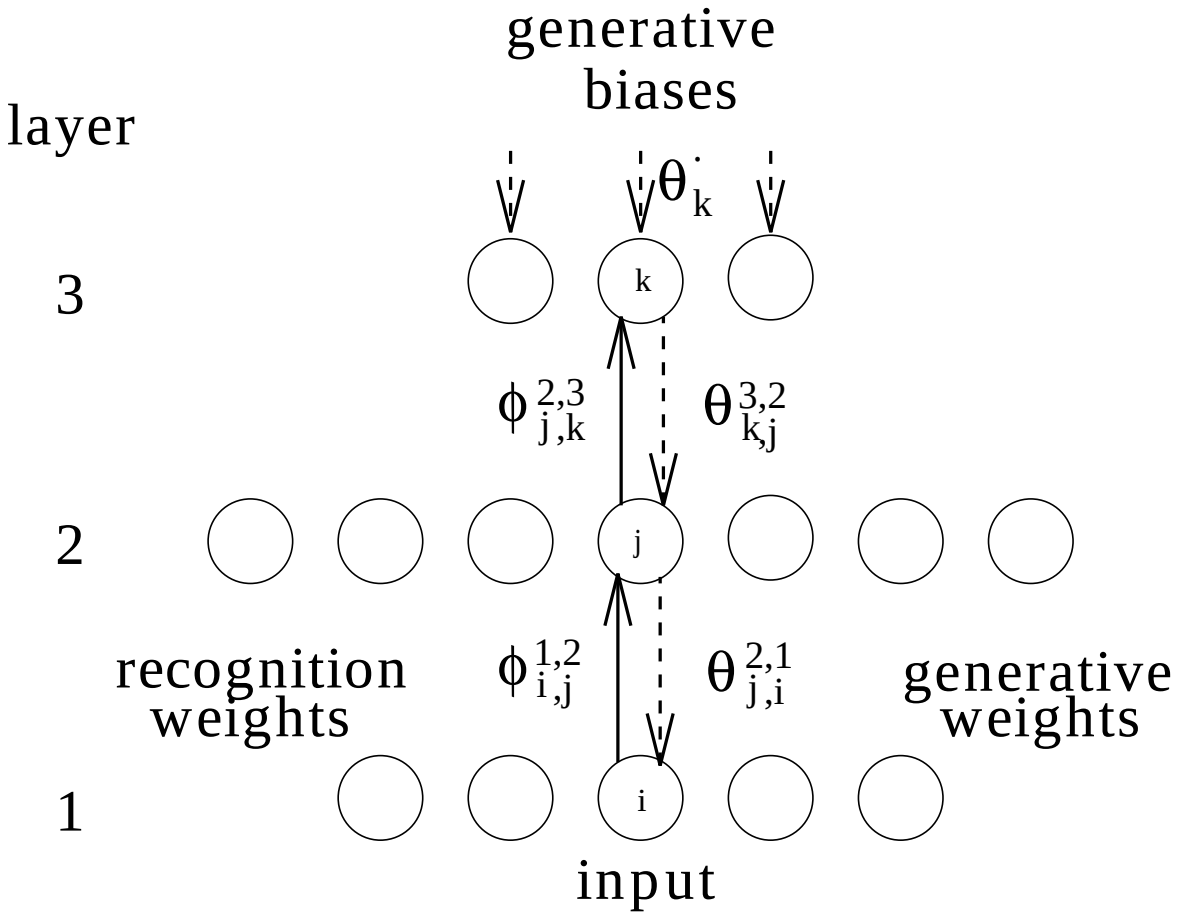
Model	IG	IGE	AUC	sAUC	NSS
Centerbias	0.00	0.0	79.6	50.0	1.22
Pixel	0.13	10.7	81.2	60.2	1.38
IttiKoch [16]	0.23	18.6	82.3	64.1	1.41
AIM [6]	0.27	22.6	82.9	65.6	1.50
eDN [48]	0.38	31.1	83.8	68.7	1.61
ICF	0.45	37.2	84.4	70.1	1.74
DeepGaze I [32]	0.56	46.1	85.8	73.0	1.92
OpenSALICON [46]	0.73	59.7	86.4	74.2	2.14
DeepGaze II	**0.98**	**80.3**	**88.3**	**77.7**	**2.48**
Gold Standard	1.22	100.0	89.9	81.2	2.82

Figure 31: From Kummerer et al. (2017) Fig. 3

IG: **情報ゲイン**, IGE: **修正情報ゲイン**, ACU: area under the ROC curve, sAUC: **シャッフル精度**, NSS: **正規化済キャンパス顕在性** normalized scanpath saliency

ヘルムホルツマシン

Dayan, Hinton, Neal, & Zemel (1995); Hinton, Dayan, Frey, & Neal (1995)



ヘルムホルツマシン

$$\log p(d|\theta) = -\sum Q_a E_a - \sum Q_a \log Q_a + \sum Q_a \log\left(\frac{Q_a}{P_a}\right) \tag{6}$$

$$= -F(d;\theta,Q) + \sum_a Q_a \log\left(\frac{Q_a}{P_a}\right) \tag{7}$$

$$q^{(l)}\left(\phi, \mathbf{s}^{(l-1)}\right) = \sigma\left(\sum s^{l-1}\phi^{(l-1,l)}\right) \tag{8}$$

$$Q_\alpha(\phi, d) = \prod\prod\left[q^{(l)}\left(\phi, \mathbf{s}^{(l-1)}\right)\right]^{s^l}\left[1 - q^{(l)}\left(\phi, \mathbf{s}^{(l-1)}\right)\right]^{1-s} \tag{9}$$

$$p_j^{(l)}\left(\theta, \mathbf{s}^{(l+1)}\right) = \sigma\left(\sum s^{(l+1)}\theta^{(l+1)}\right) \tag{10}$$

$$p(\alpha|\theta) = \prod\prod\left[p_j^{(l)}\left(\theta, \mathbf{s}^{(l+1)}\right)\right] \tag{11}$$

モデル: ヘルムホルツマシン

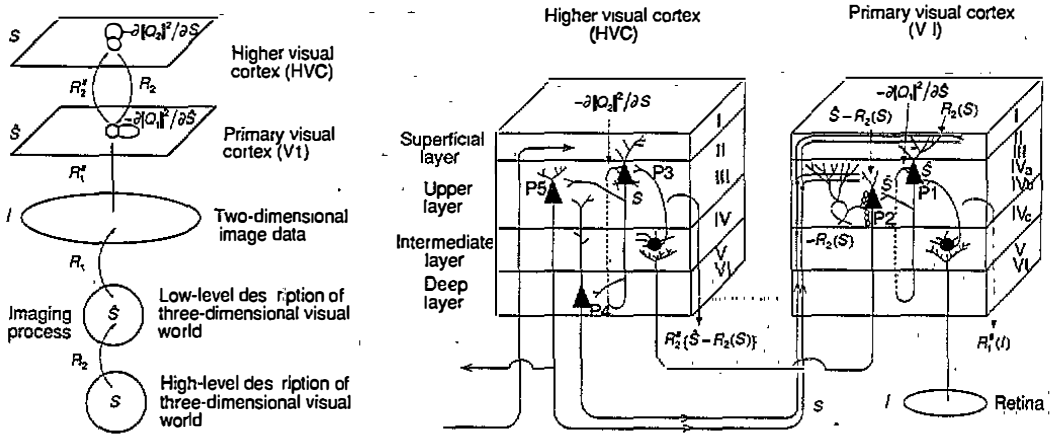


Figure 32: From Kawato et al. (1993) Fig. 1 より

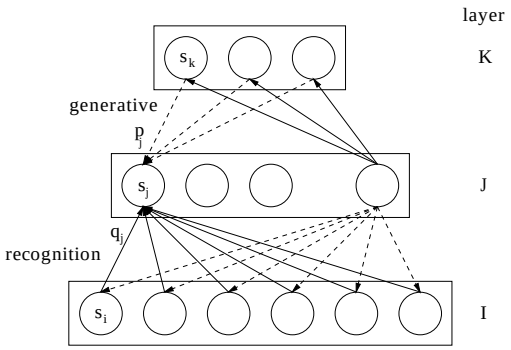


Figure 33: From Hinton et al. (1995) Fig. 1 より

- 上位層は下位層からの情報をサンプリング → 認識形成 トップダウン
- 下位層は上位層からの情報を受けとる → 情報再構成 ボトムアップ

ボトムアップ処理による認識とトップダウン処理による (こう見えるはずだという思い込みの) 生成を n $(n = 2, \ldots, 4)$ 回繰り返す → パレイドリア成立

定式化

思い込みの印象 α と入力画像 d を用いて

$$C(\alpha, d) = C(\alpha) + C(d|\alpha) \tag{12}$$

$$= \sum_{\ell \in L} \sum_{j \in \ell} C(s_j^\alpha) + \sum_i C(s_i^d|\alpha) \tag{13}$$

上式を用いて結合係数の更新を行う

$$\Delta w_{kj} = \epsilon s_k^\alpha \left(s_j^\alpha - p_j^\alpha \right), \tag{14}$$

$$C(d) = \sum_\alpha Q(\alpha|d) C(\alpha, d) - \left[- \sum_\alpha Q(\alpha|d) \log Q(\alpha|d) \right]. \tag{15}$$

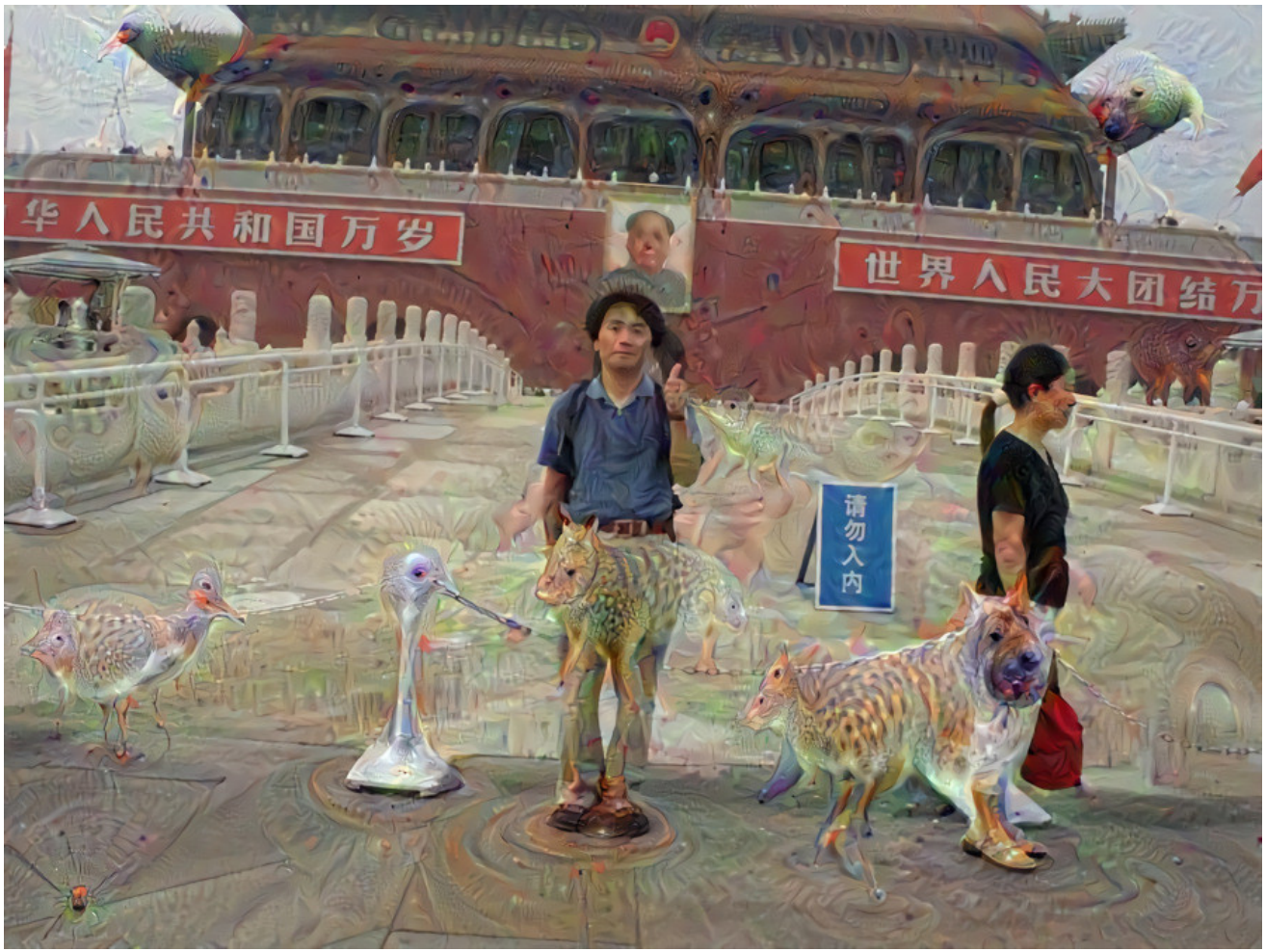
$$p\left(\alpha|d\right) = \frac{e^{-C(\alpha,d)}}{\sum_\beta e^{-C(\beta,d)}} \tag{16}$$

$$\Delta s_{j,t+1} = \epsilon s_{j,t}^\gamma (s_{j,t}^\gamma - q_{j,t}^\gamma) \tag{17}$$

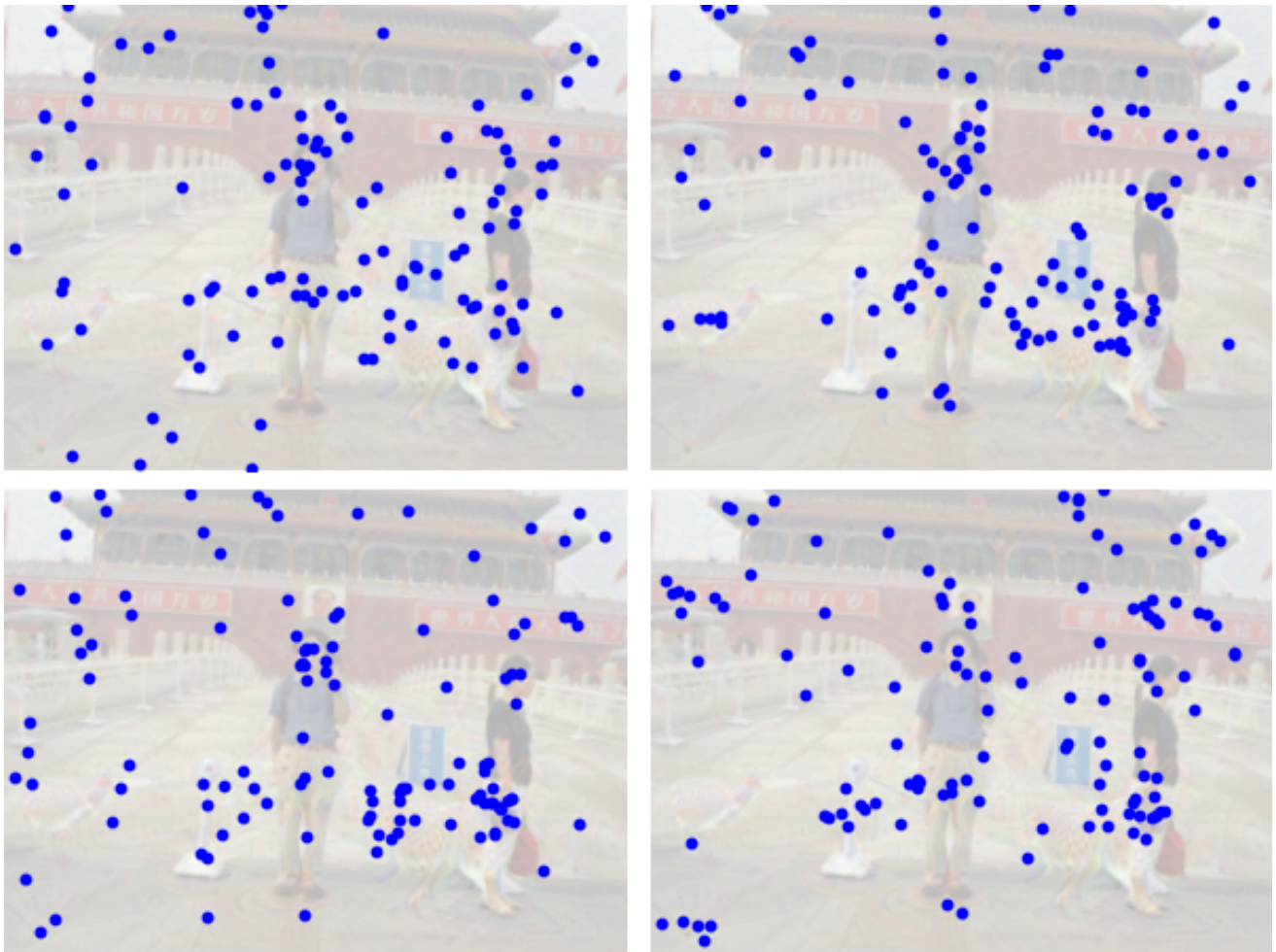
全体の良い表象が得られるまで，すなわち下位層の活性を再構築するように複数回繰り返す

計算例

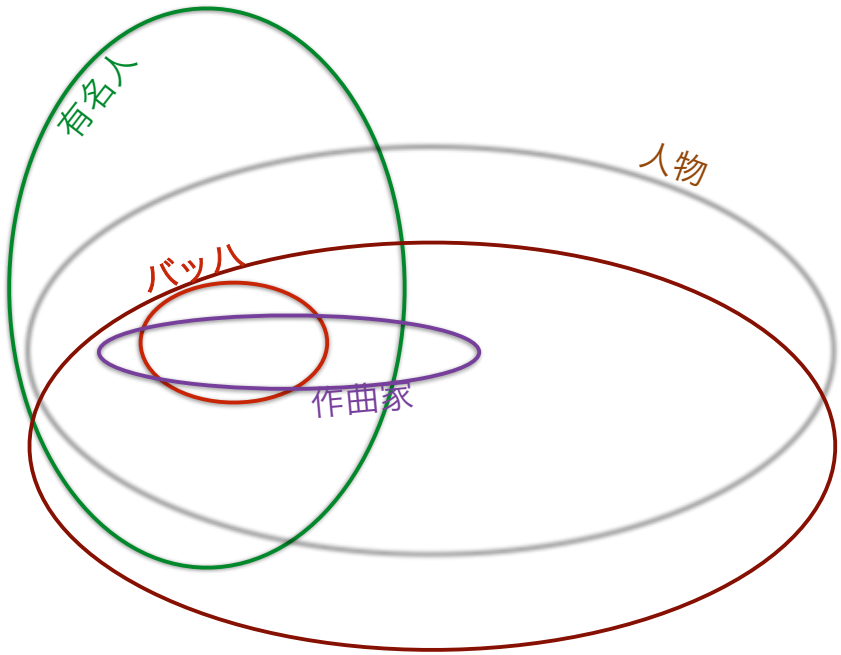
計算例



計算例 (2) 眼球運動のサンプリング



意味の広がりと非対称性



For 単語 i:

$$\mathbb{N}(x;\mu_i;\Sigma_i) \propto -\underbrace{\log\det(\Sigma_i)}_{\text{loss}} - \overbrace{(\mu_i - x)^\top \Sigma_i^{-1} (\mu_i - x)}^{\text{Maharanobis distance}} \tag{18}$$

$$E\left(\text{word}_i, \text{word}_i\right) = \int_x \mathbb{N}\left(x; \mu_i, \Sigma_i\right) \mathbb{N}\left(x; \mu_j, \Sigma_j\right) \, dx \tag{19}$$

$$= \mathbb{N}\left(0; \mu_i - \mu_j, \Sigma_i + \Sigma_j\right) \tag{20}$$

$$\propto -\log\det\left(\Sigma_i + \Sigma_j\right) - \left(\mu_i - \mu_j\right)^\top \left(\Sigma_i + \Sigma_j\right)^{-1} \left(\mu_i - \mu_j\right) \tag{21}$$

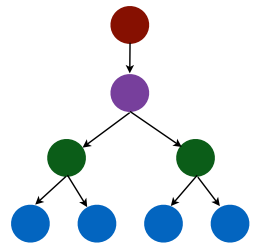
Gaussian Embedding

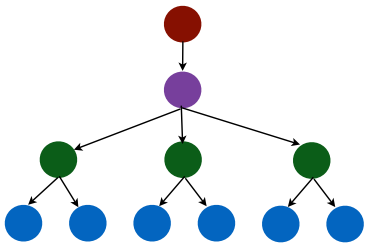
. . . German musician and composer of the Baroque . . .

$$E\,(\text{composer}, \text{musician}) > E\,(\text{composer}, \text{banana}) \tag{22}$$

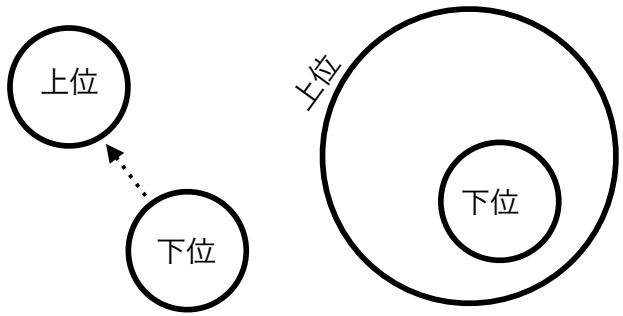
$$\max\left(0, m + \text{KL}\left(C_{pos} \,\|w)\right) - \text{KL}\left(C_{neg} \,\|w\right)\right. \tag{23}$$

概念の階層



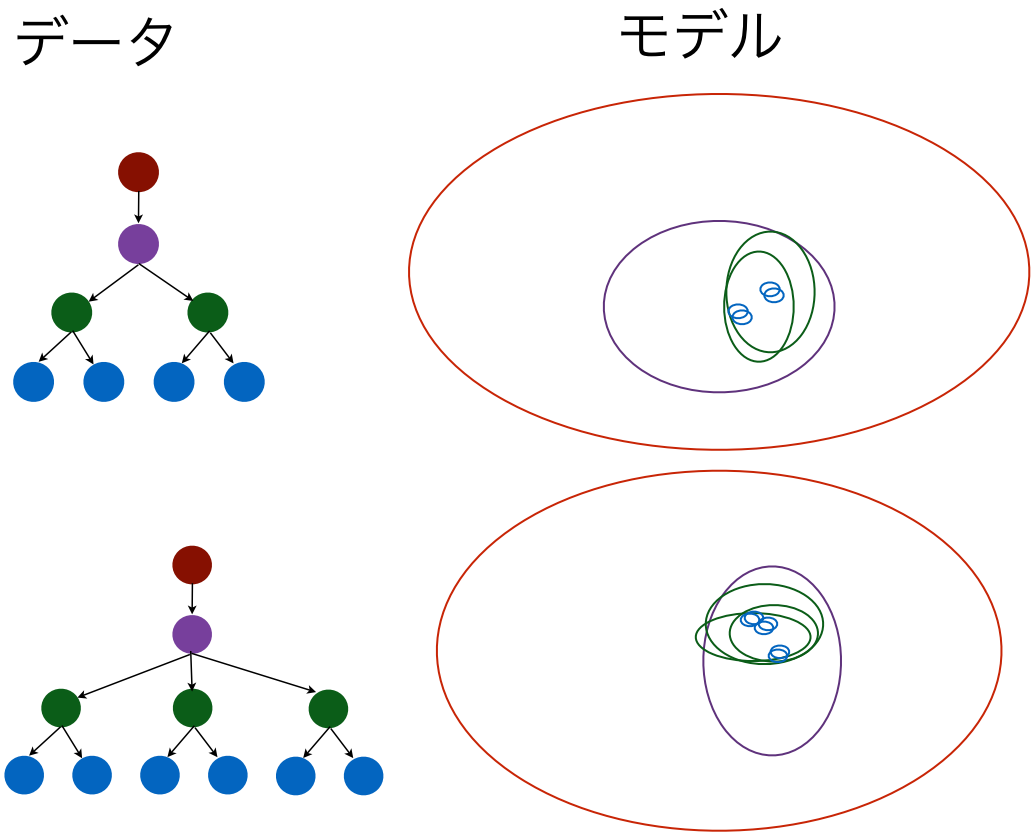


下位概念 $\vdash$ 上位概念

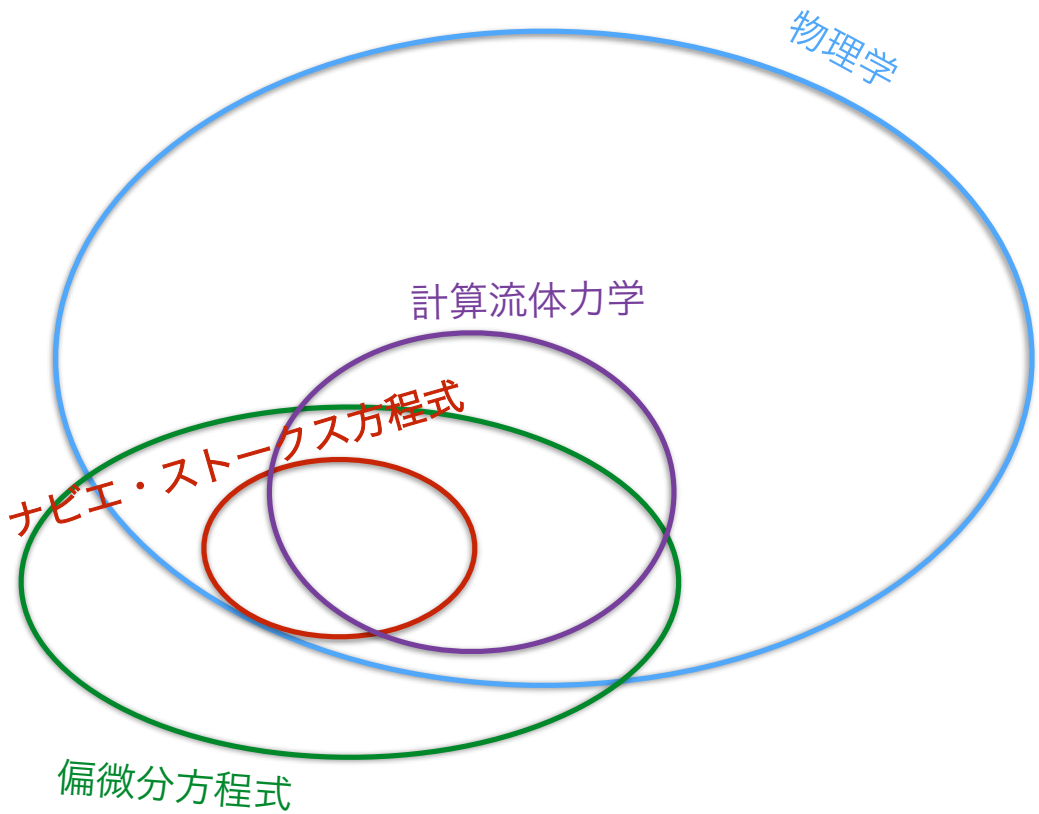


$$\mathrm{KL}\left(v_{\text{下位概念}} \middle\| v_{\text{上位概念}}\right)$$

概念の階層



計算例



これって Matsuka, Sakamoto, & Chouchourelou (2008); Vilnis & McCallum (2015) **と同じじゃね？**

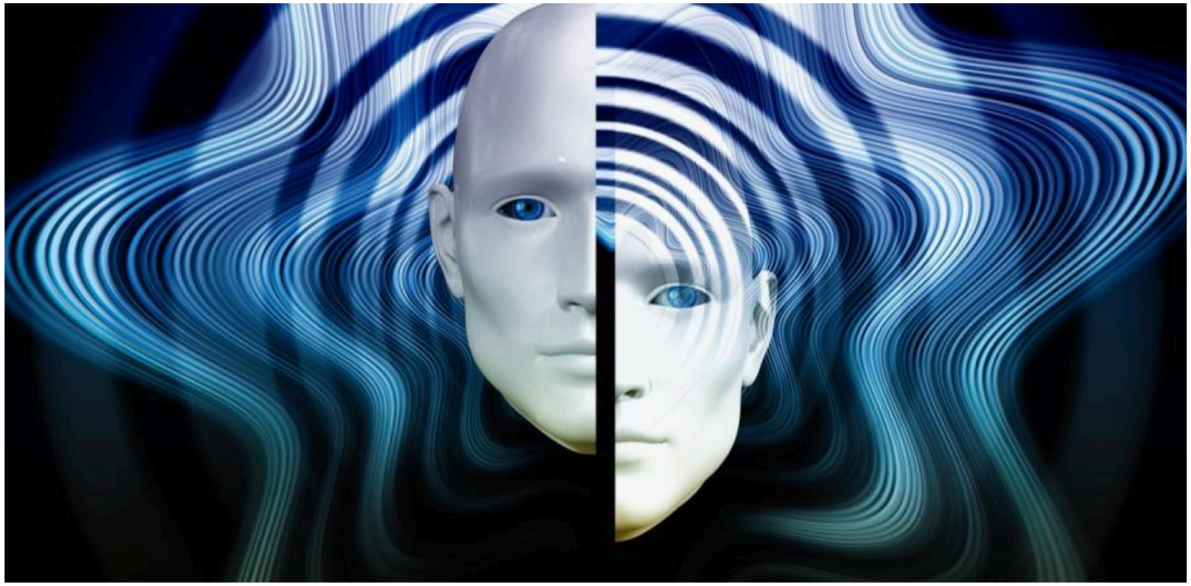
謝辞

本発表に際しまして特別なご配慮いただきました先生方に感謝申し上げます

- 貴重な意見と資料を提供していただいた 松沢病院 西尾慶之 先生
- 発表についてご配慮をいただいた本大会プログラム委員長 森田純哉 先生

おまけ

シンギュラリティサロン https://singularity37.peatix.com/, 9 月 22 日 @ 大阪, 10 月 26 日 @ 東京



シンギュラリティサロン #37　「自然言語処理と画像処理における最近の注意モデル」浅川 伸一

DESCRIPTION

講演

13:30-15:00　講演:「自然言語処理と画像処理における最近の注意モデル」

15:00-15:30　自由討論

講師

浅川伸一　（東京女子大学 情報処理センター）

定員　100名（先着順・入場料無料）

Sun Sep 22, 2019
1:30 PM - 3:30 PM JST
Add to Calendar

VENUE グランフロント大阪・ナレッジサロン・プレゼンラウンジ

TICKETS

一般（無料）

GET TICKET

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In Y. Bengio & Y. LeCun (Eds.), *Proceedings in the International Conference on Learning Representations (ICLR).* San Diego, CA, USA.

Bloom, F. E., & Lazerson, A. (1988). *Brain, mind, and behavior* (2nd ed.). New York, NY: Freeman.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *35*, 185-207.

Broadbent, D. E. (1958). *Perception and communication*. Oxford,UK: Pergamon.

Buschman, T. J., & Kastner, S. (2015). From behavior to neural dynamics: An integrated theory of attention. *Neuron*, *88*, 127–144.

Crick, F. (1984). Function of the thalamic reticular complex: the search light hypothesis. *Proceedings of the National Academy of Sciences*, *81*, 4586–4590.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. *Neural Computation*, *7*, 889–904.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 578–588.

Eriksen, C. W., & St.James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, *40*, 225–240.

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. *Current Biology*, *23*, 2553-2558.

Heilman, K. M., & Valenstein, E. (1979). Mechanisms underlying hemispatial neglect. *The Annals of Neurology*, *5*, 166-170.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The "wake-sleep" algorithm for unsupervised neural networks. *Science*, *268*, 1158–1161.

Itti, L., & Borji, A. (2014). Computational models: Bottom-up and top-down aspects. In A. C. Nobre & S. Kastner (Eds.), *The oxford handbook of attention* (p. 1122-1158). Oxford University Press.

Itti, L., & Borji, A. (2015). Computational models of attention. *arXiv preprint*.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 1–11.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254-1259.

Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Boston, MA, USA.

Kawato, M., Hayakawa, H., & Inui, T. (1993). A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network: Computation in Neural Systems*, *4*, 415–422.

Kimura, A., Yonetani, R., & Hirayama, T. (2013). Computational models of human visual attention and their implementations: A survey. *IEICE Transactions of Information & Systems*, *E96-D*, 562-578.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Revivew of Neuroscience*, *30*, 57–78.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219-227.

Krauzlis, R. J., Lovejoy, L. P., & Zénon, A. (2013). Superior colliculus and visual spatial attention. *Annual Review of Neuroscience*, *36*.

Kummerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding low- and high-level contributions to fixation prediction. In *The IEEE International Conference on Computer Vision (ICCV)* (pp. 4789–4798). Venice, Italy.

Liu, X., He, P., Chen, W., & Gao, J. (2019). Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4487–4496). Florence, Italy: Association for Computational Linguistics.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint*.

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint*, *cs.CL*, 1508.04025.

Matsuka, T., Sakamoto, Y., & Chouchourelou, A. (2008). Modeling a flexible representation machinery of human concept learning. *Neural Networks 21 (2008) 289–302*, *21*, 289-302.

Milanese, R., Wechsler, H., Gill, S., Bost, J.-M., & Pun, T. (1994). Integration of bottom-up integration of bottom-up and top-down cues for visual attention using non-linear relaxation. In *The proceedings of CVPR, IEEE – Institute of Electrical and Electronics Engineers* (p. 781-785). Dallas Texas, USA.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Revivew of Neuroscience*, *24*.

Monosov, I. E., & Thompson, K. G. (2009). Frontal eye field activity enhances object identification during covert visual search. *Journal of Neurophysiology*, *102*, 3656–3672.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3-25.
Sperry, R. W. (1961). Cerebral organization and behavior. *Science*, *133*, 1749–1757.
Sperry, R. W. (1968). Hemisphere deconnection and unity in conscious awareness. *American Psychologist*, *28*, 723–733.
Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger (Eds.), *Advances in Neural Information Processing Systems (NIPS)* (Vol. 27, pp. 3104–3112). Montreal, BC, Canada.
Treisman, A. (1964). Selective attention in man. *British Medical Bulletin*, *20*, 12-16.
Treisman, A. (1988). Feature and objects: The fourteenth bartlett memorial lecture. *The quarterly Journal of Experimental Psychology*, *40A*, 201–237.
Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97-136.
Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *JEP:General*, *114(3)*, 285-310.
Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*, 282–299.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Kaiser, Ł. (2017). Attention is all you need. *arXiv preprint*.
Vilnis, L., & McCallum, A. (2015). Word representations via gaussian embedding. In Y. Bengio & Y. LeCun (Eds.), *The proceedings of International Conference on Learning Representations (ICLR).* San Diego, CA, USA.
Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In *Computer Vision and Pattern Recognition (CVPR).* Boston, MA, USA.
Wang, W., & Shen, J. (2018). Deep visual attention prediction. *IEEE Transactions on Image Processing*, *27*, 2368-2378.
Wardak, C., Olivier, E., & Duhamel, J.-R. (2004). A deficit in covert attention after parietal cortex inactivation in the monkey. *Neuron*, *42*, 501–508.
Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202-238.
Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R. S., & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. *arXiv:1502.03044*.