Learning to Draw

Emergent Communication through Sketching

Daniela Mihai and Jonathon Hare
{adm1g15, jsh2}@soton.ac.uk

Vision, Learning and Control Group
School of Electronics and Computer Science
University of Southampton

Emergent Communication

- Emergent communication is the study of how agents *learn to utilise their communication channel to convey information to solve a task.*
- Historically, most literature has focussed on token-based communication (e.g. modelling written language).

Emergent Communication

- Emergent communication is the study of how agents *learn to utilise their communication channel to convey information to solve a task.*
- Historically, most literature has focussed on token-based communication (e.g. modelling written language).
- Referential games are often used as a playground.

Referential Communication Games

A Referential Game*: Alice must communicate to Bob which image she has (Bob has that image, plus many distractors). Communication is one-way only. Alice knows nothing about the distractors Bob has (they could all be white boats!).

* David K. Lewis. Convention: A Philosophical Study. Wiley-Blackwell, 1969.

Emergent Communication

- Emergent communication is the study of how agents *learn to utilise their communication channel to convey information to solve a task.*
- Historically, most literature has focussed on token-based communication (e.g. modelling written language).
- Referential games are often used as a playground.
- **We seek to instead look at visual communication channels in referential games.**

Emergent Communication

Challenges and Questions (I)

- Understanding “what” is being communicated is *hard.*
    - **Could a *constrained visual communication channel* be more interpretable?**

Emergent Communication

Challenges and Questions (II)

- Training of agents is sensitive to “hashing” solutions whereby communication is achieved in a way that relies on *non-semantic* features, or features that a human wouldn’t or couldn’t use.
  - **What *inductive biases* in the model and during training are needed to stop this happening?**

Emergent Communication

Challenges and Questions (III)

- **Can we achieve successful Agent-Human communication with a model trained with inter-Agent self-supervised learning?**

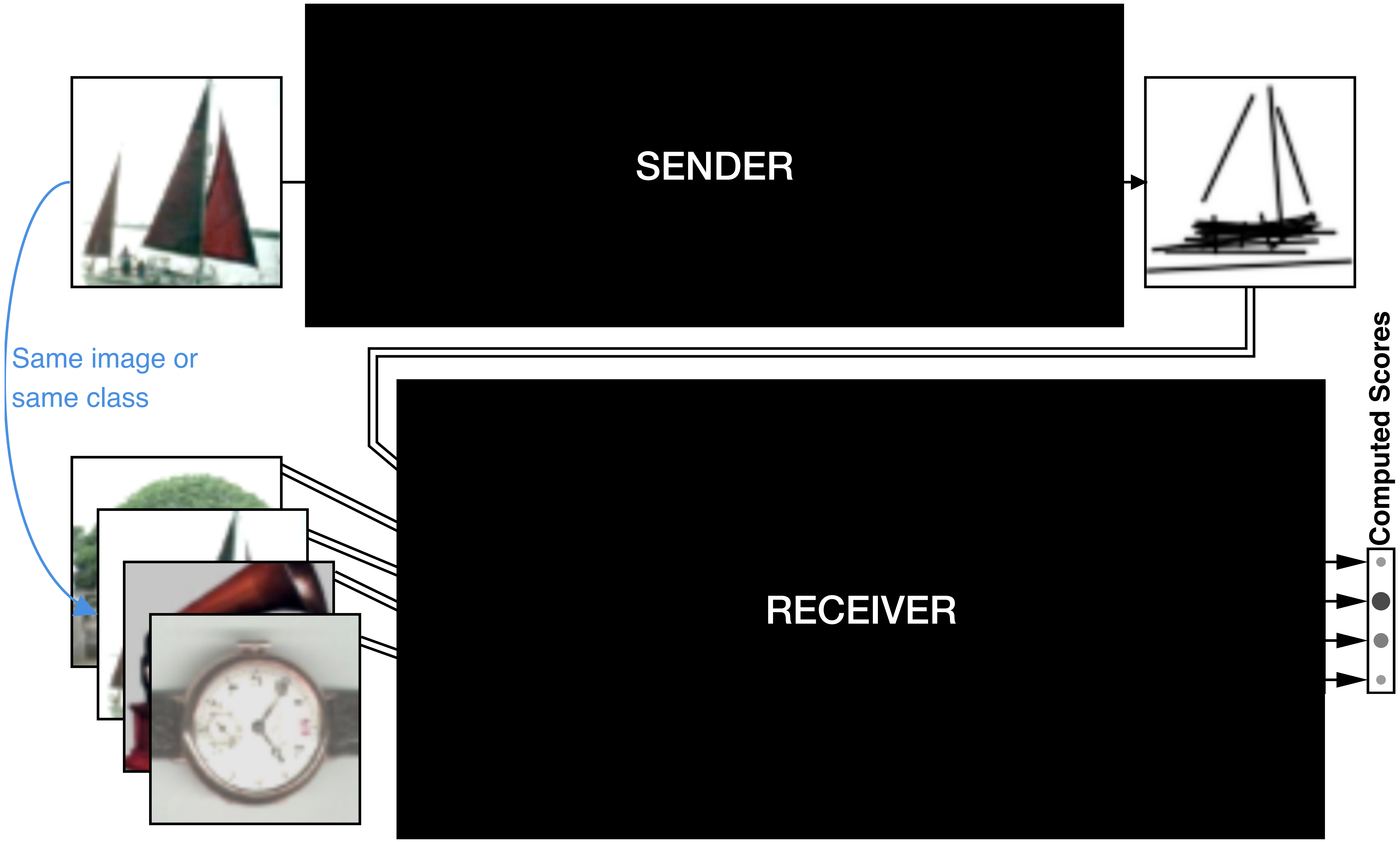
A model for learning to communicate by drawing

UNIVERSITY OF Southampton

A model for learning to communicate by drawing

The Game Environment

Objective is for receiver to correctly guess sender's image amongst distractors.



A model for learning to communicate by drawing

The Game Environment

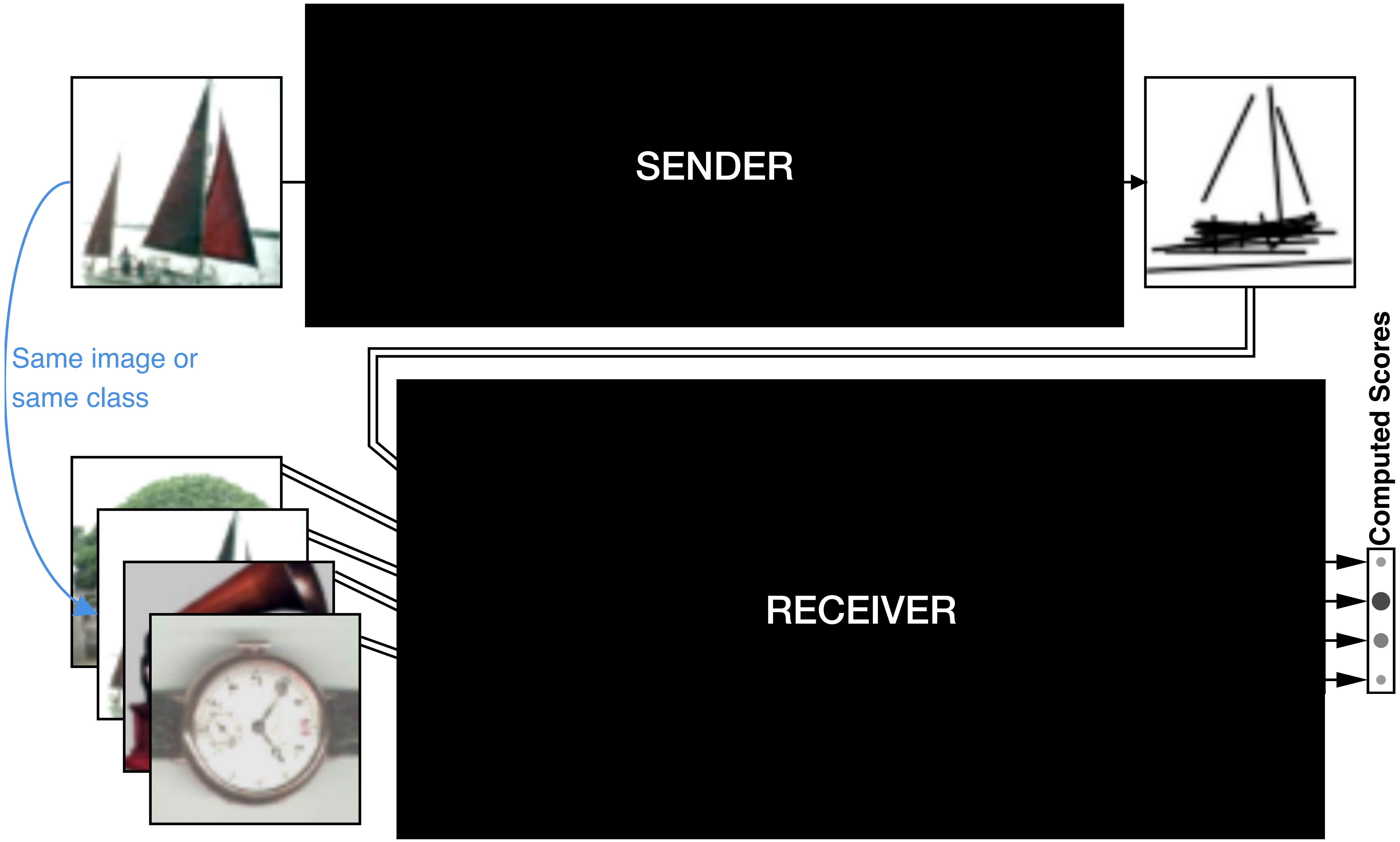
Three game variants:

Original: receiver's images are 99 randomly sampled distractors + target

Object-Oriented: receiver's images are from different classes.

In **same** target matches sender.

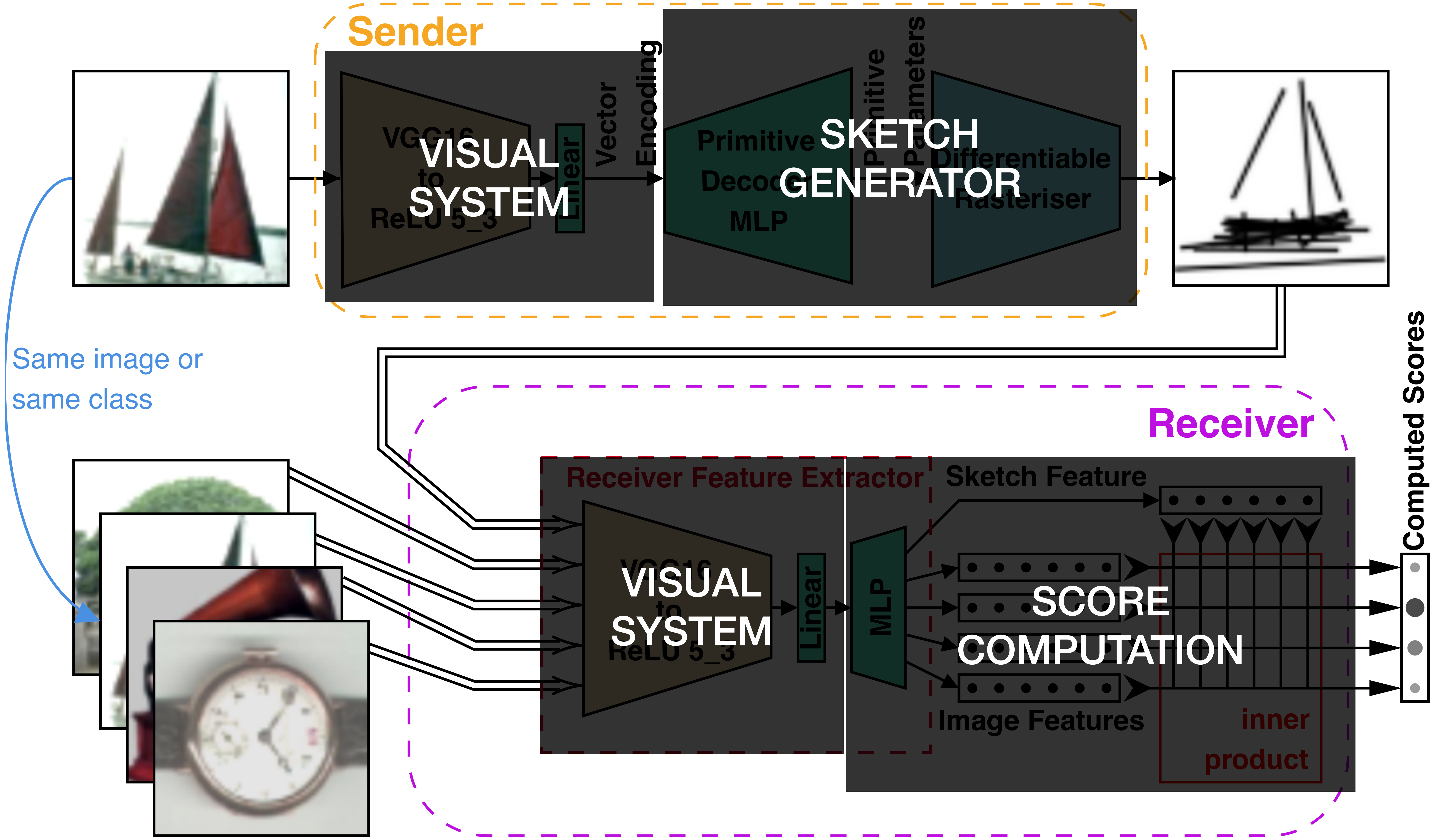
In **different** target matches class of sender's image.



A model for learning to communicate by drawing

The Agents' Architecture: Overview

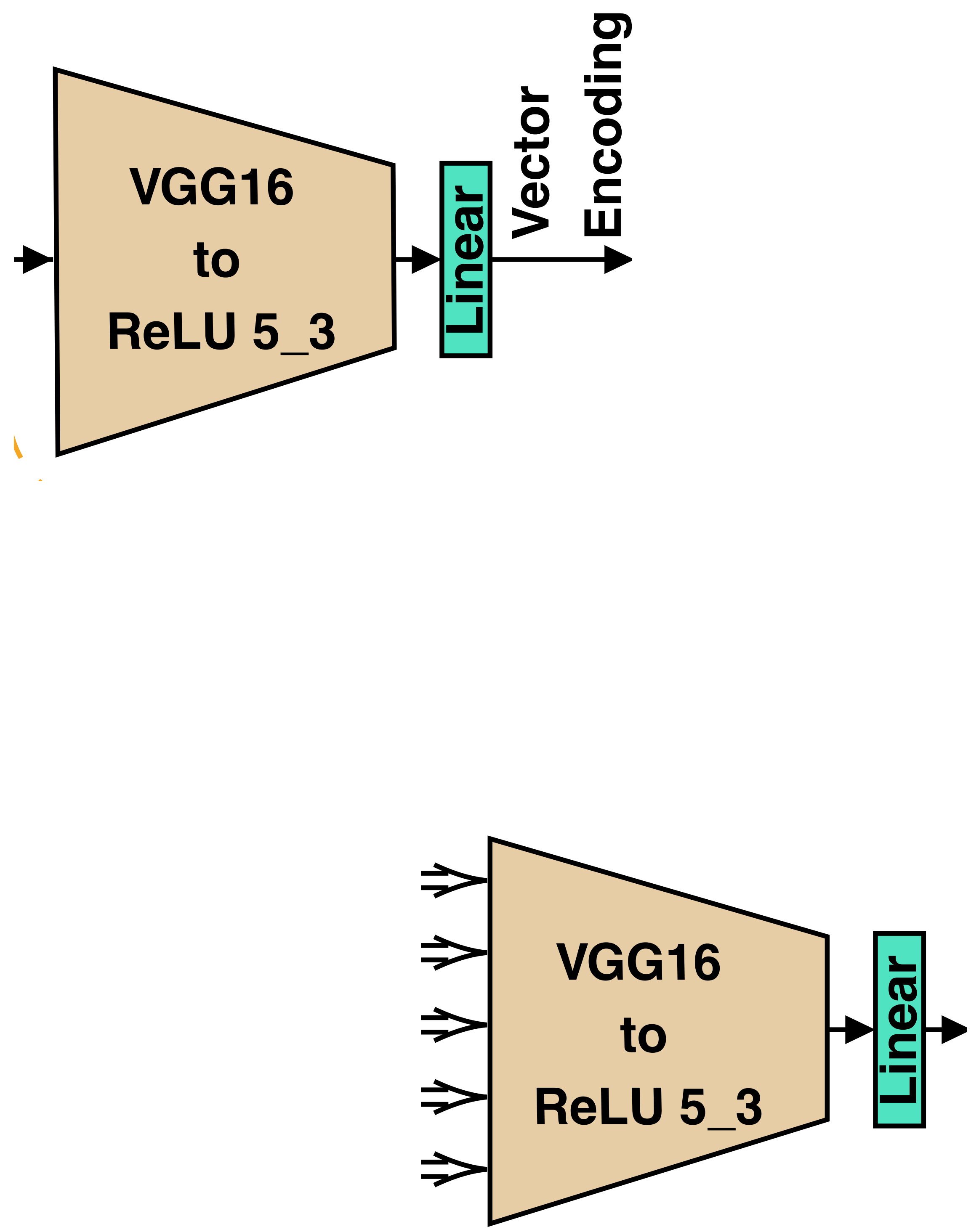
Agents consist of a visual system plus a task-specific module.



A model for learning to communicate by drawing

The Agents' Architecture: Visual System

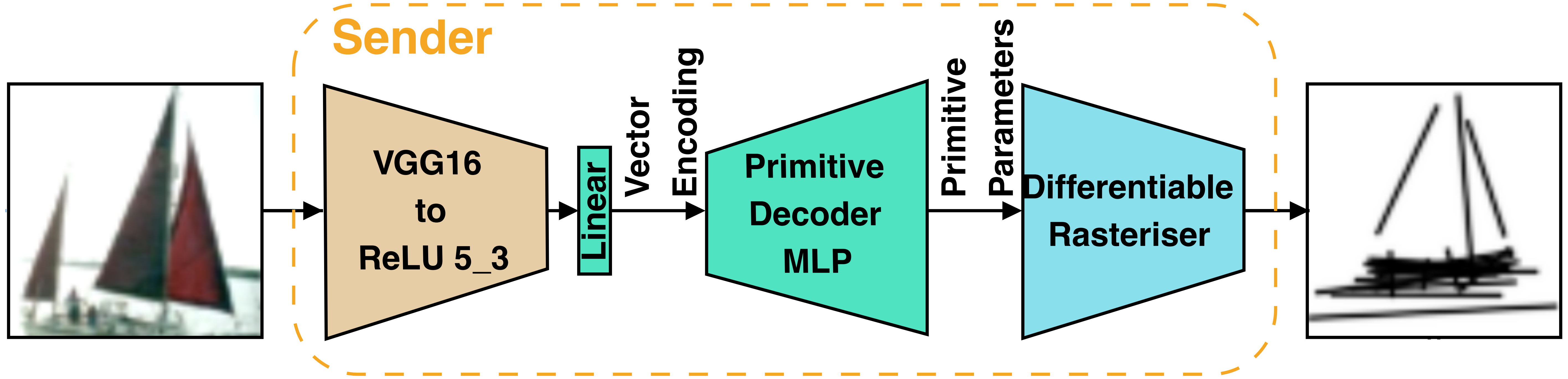
The visual system is a VGG16 with **fixed pretrained weights** from ImageNet or StylizedImageNet followed by a learned linear projection.



A model for learning to communicate by drawing

The Agents' Architecture: Sender agent

The sender agent encodes the input with the visual system and predicts the start and end points of a set of lines and renders these into an image.



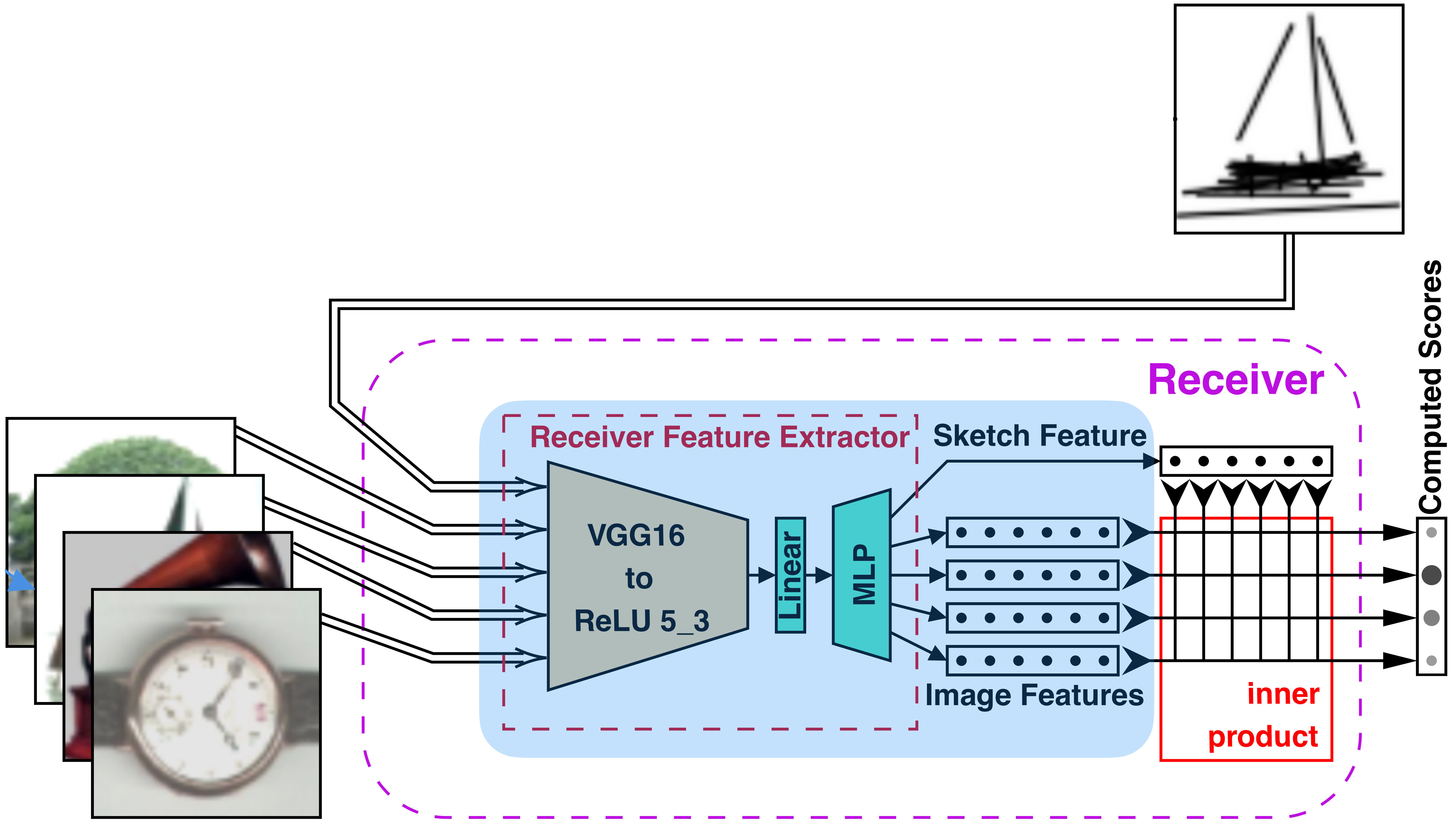
We developed a differentiable rasteriser* that allows gradients to flow between the resultant raster and the line parameters.

* Daniela Mihai and Jonathon Hare. "Differentiable Drawing and Sketching." *arXiv preprint arXiv:2103.16194* (2021).

A model for learning to communicate by drawing

The Agents' Architecture: Receiver agent

The receiver agent encodes each of its inputs with the visual system, and projects them into a learned space of features with an MLP.

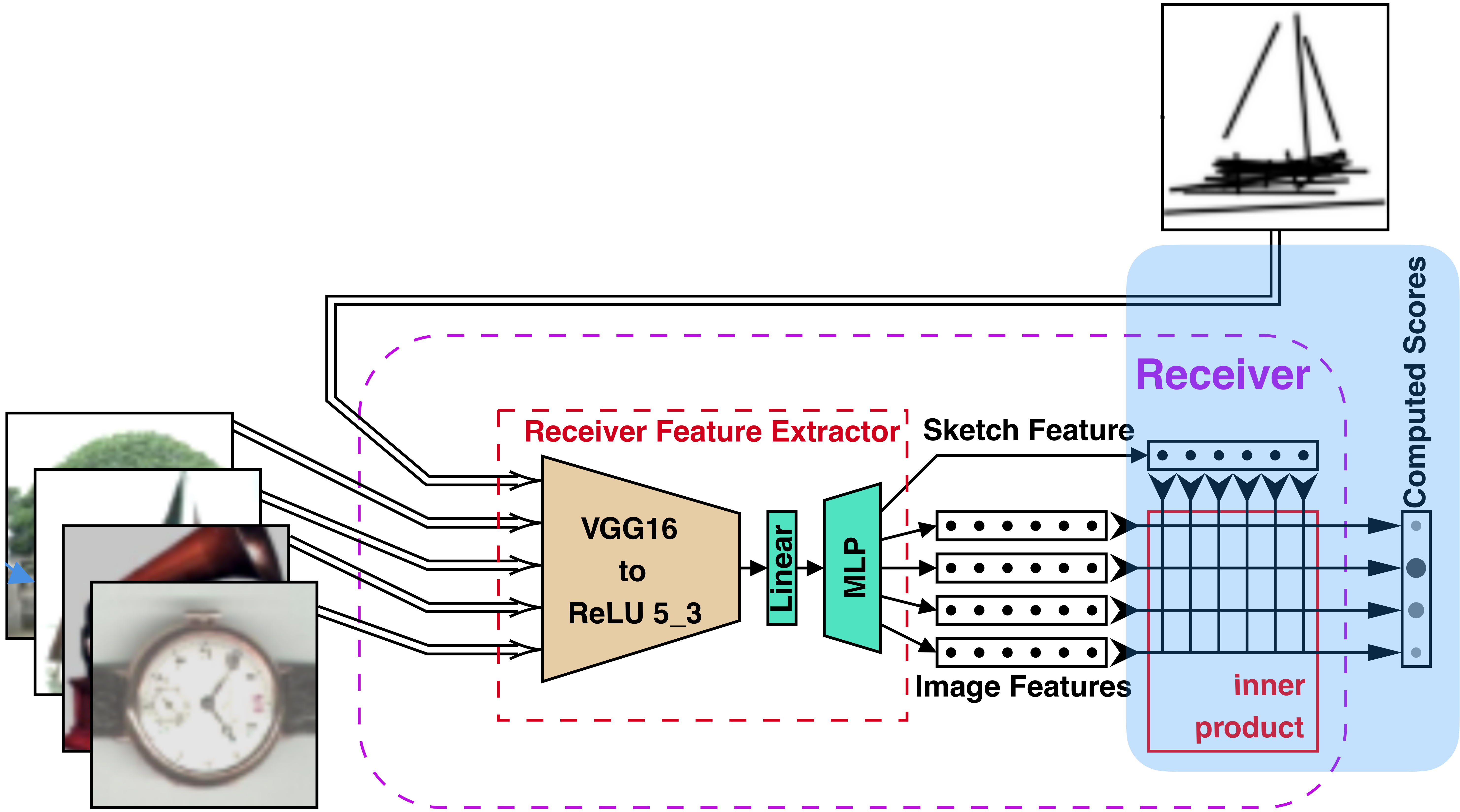


University of Southampton

A model for learning to communicate by drawing

The Agents' Architecture: Receiver agent

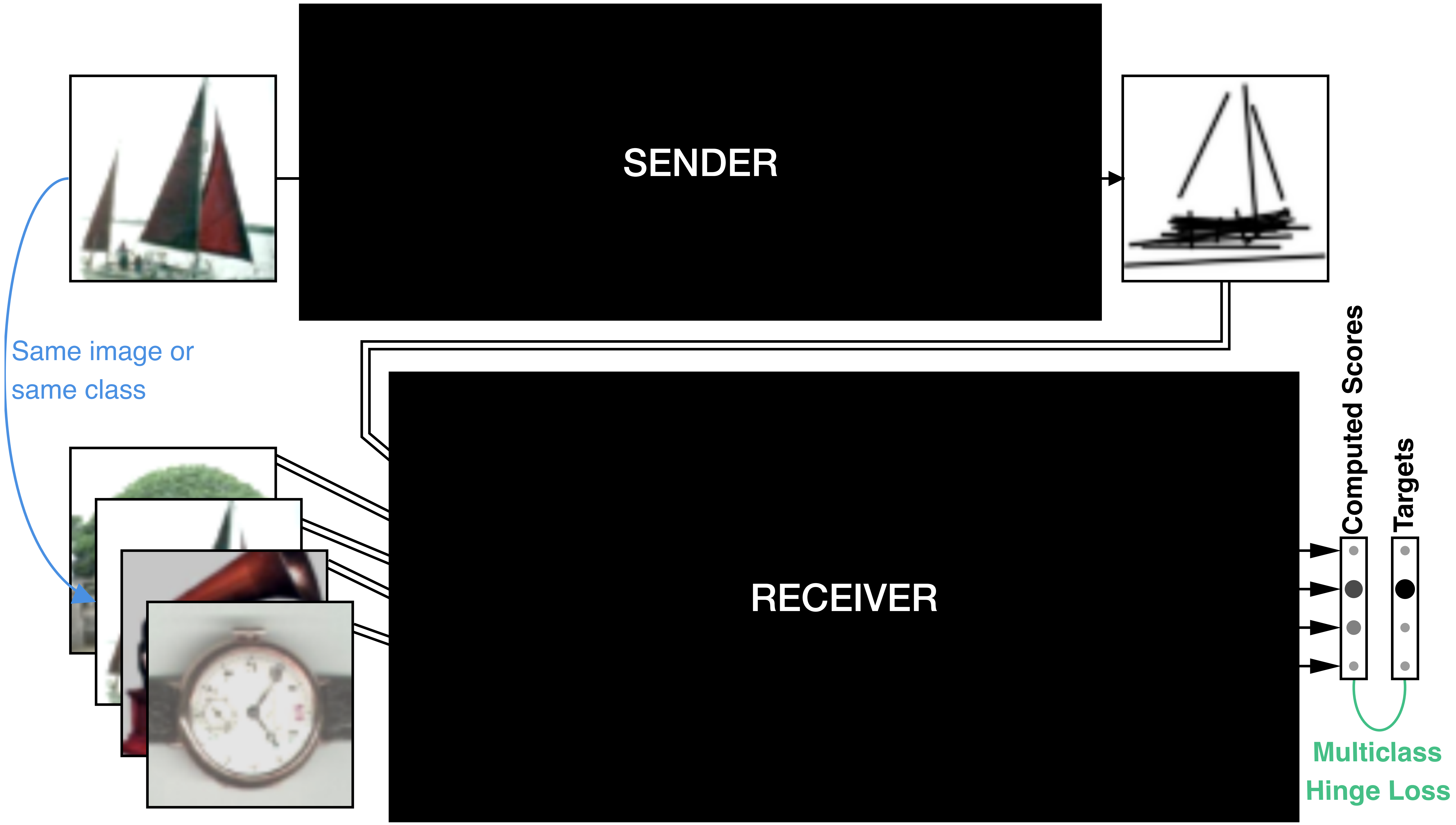
The agent uses the inner product between the sketch feature and each image feature to compute a score for each image.



A model for learning to communicate by drawing

Training

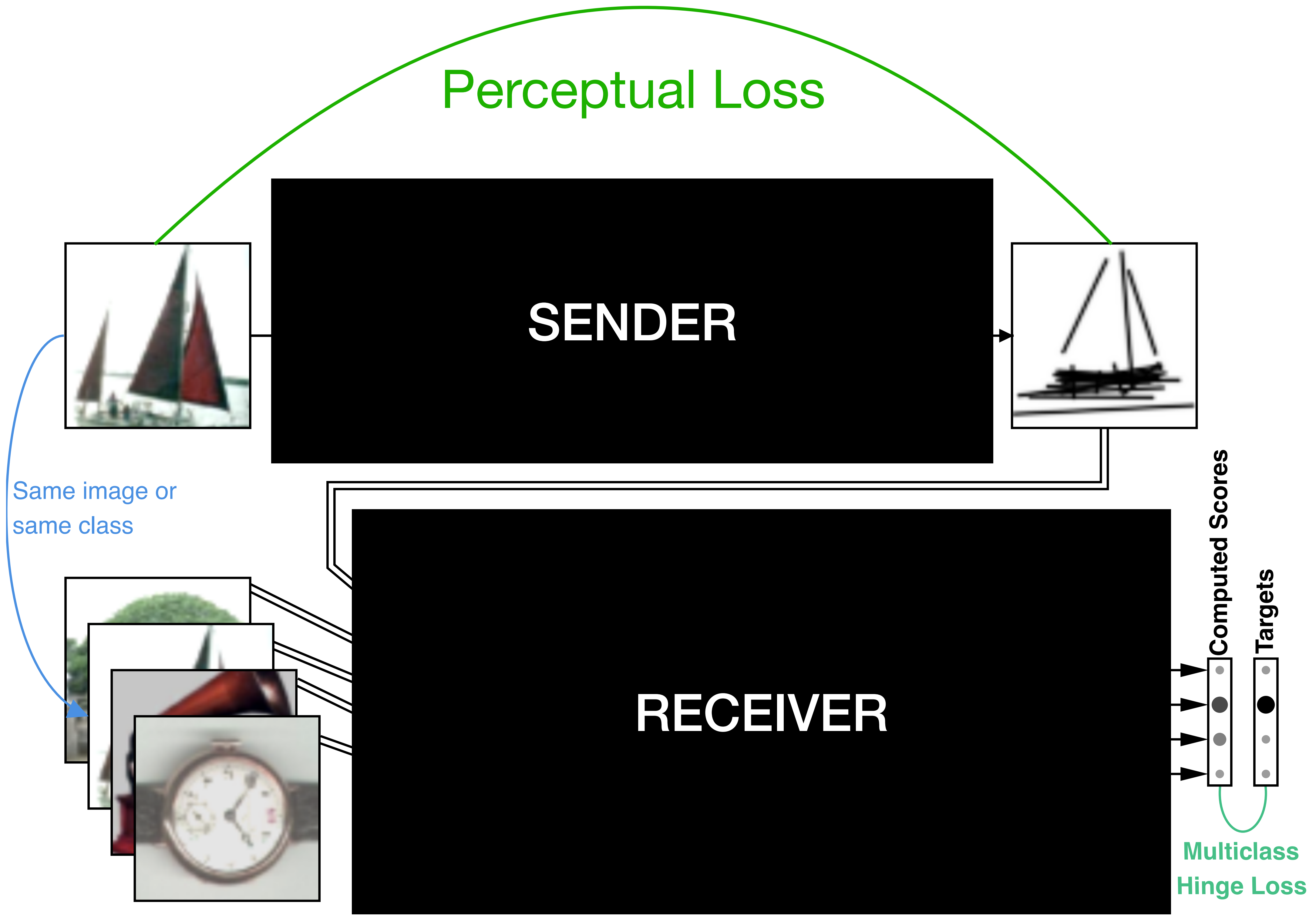
A **multiclass hinge loss*** is used with a gradient-based optimiser (Adam) to learn the parameters of both agents.



* Other losses available: cross-entropy works well too

A model for learning to communicate by drawing

Making the sender's sketches more perceptually relevant

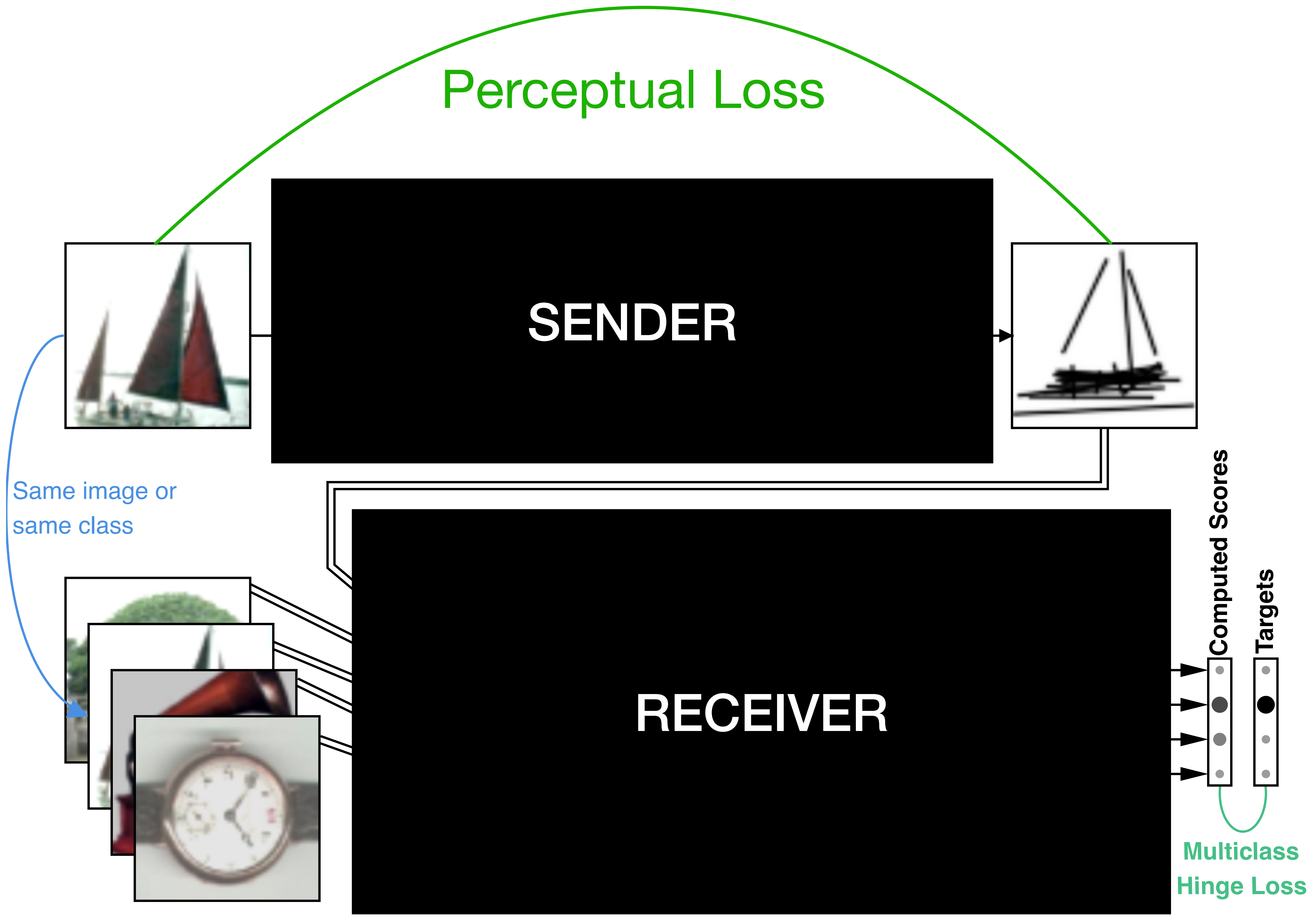


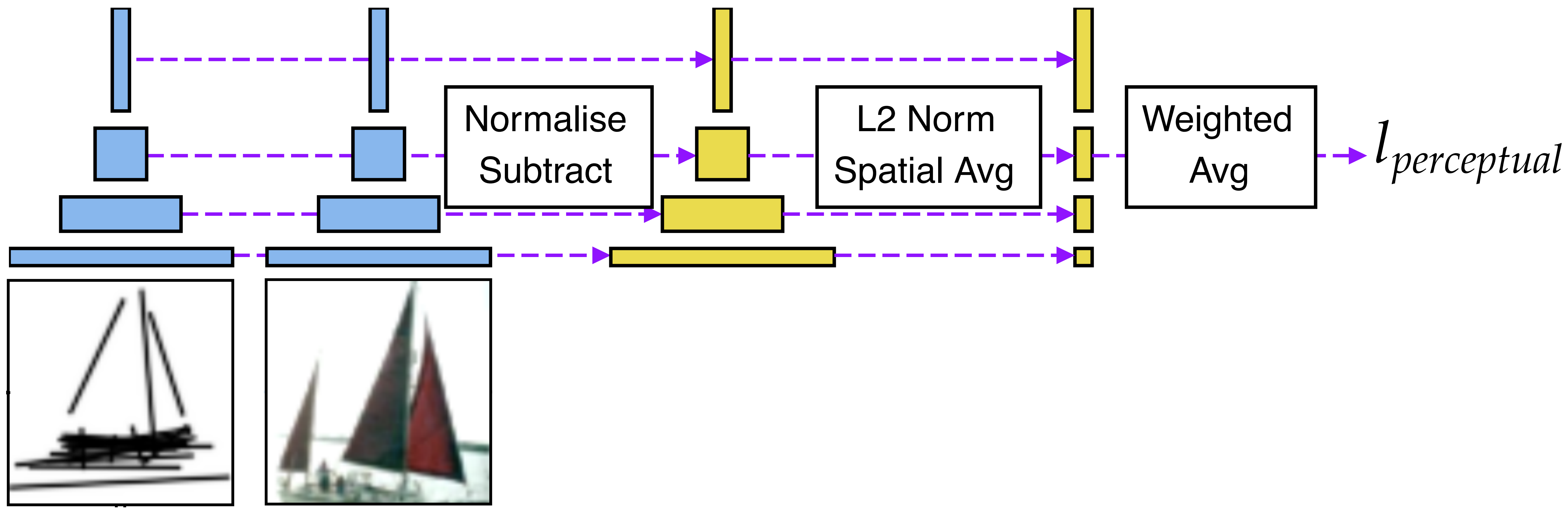
The sketches created by the sender will often look **random**.

Incorporating a **perceptual loss** will be shown to help.

A model for learning to communicate by drawing

Making the sender's sketches more perceptually relevant





We experiment with a simple* perceptual loss computed across the layers of internal representation of the VGG16-based visual system.

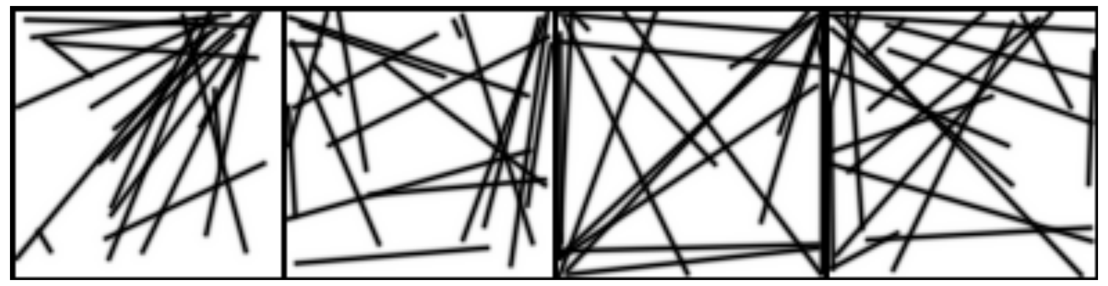
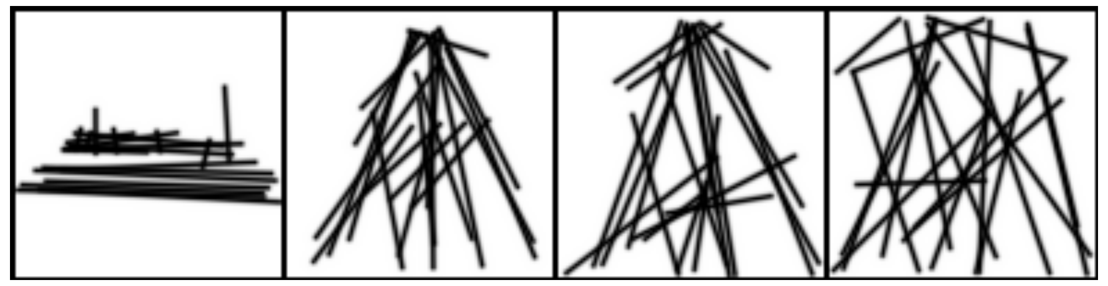
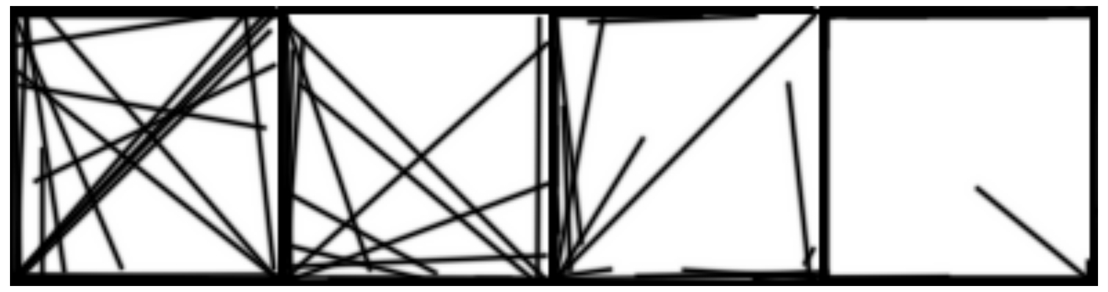
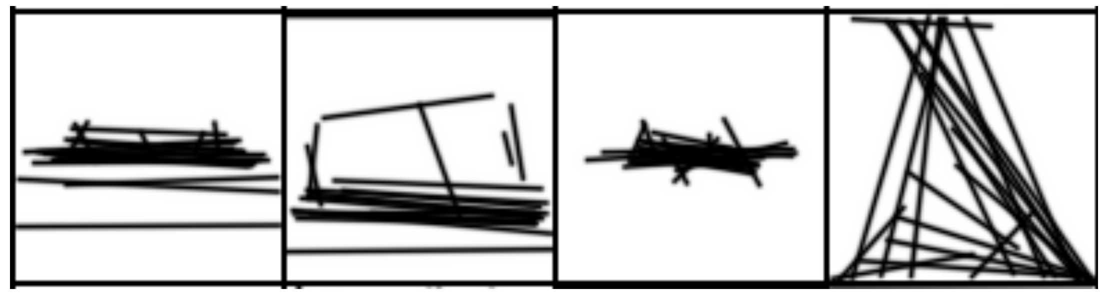
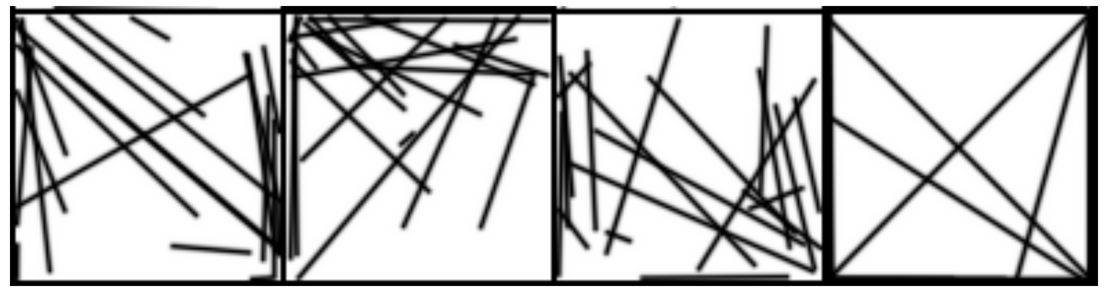
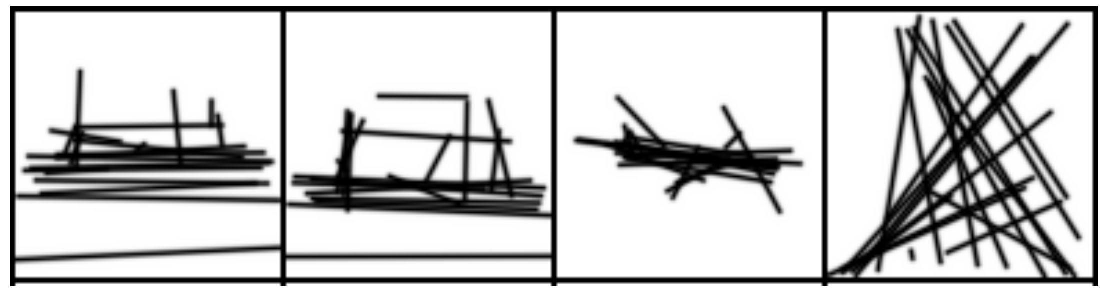
* Inspired by: Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. "The unreasonable effectiveness of deep features as a perceptual metric." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586-595. 2018.

Experiments

UNIVERSITY OF Southampton

Experiments

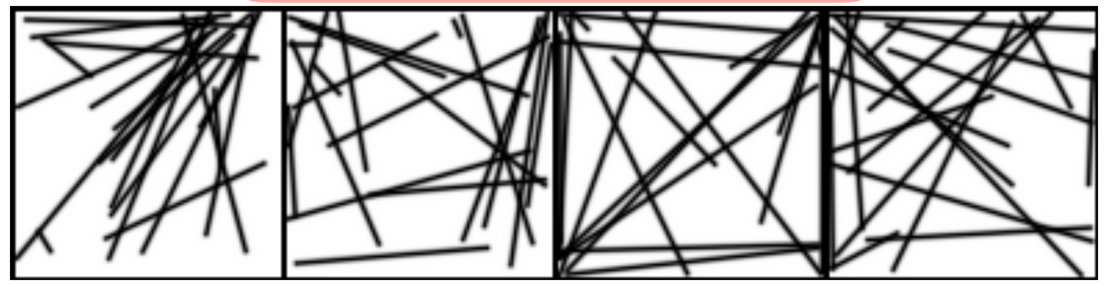
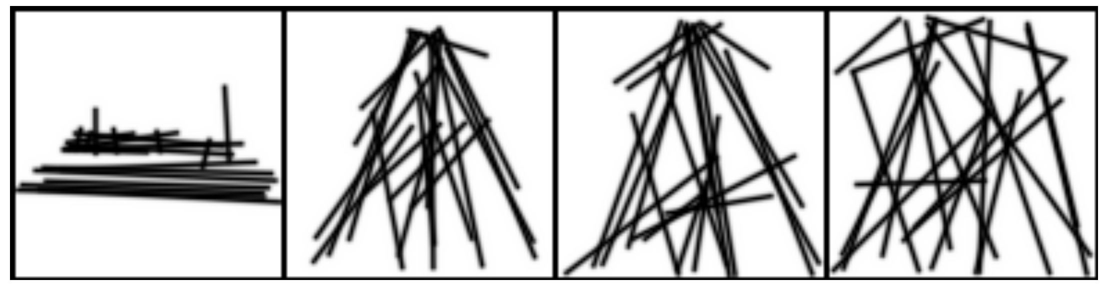
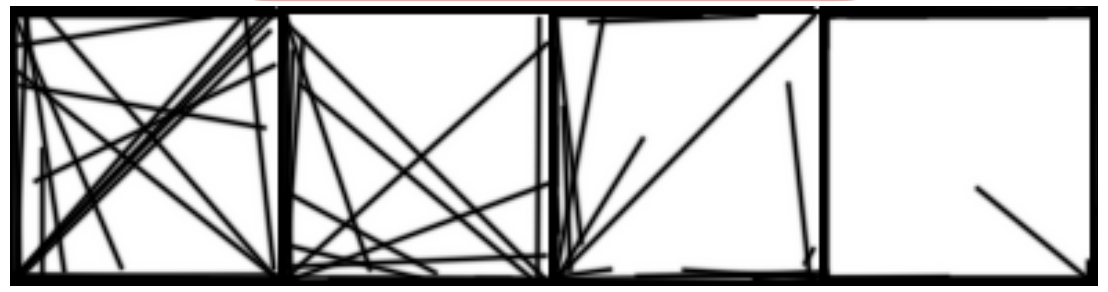
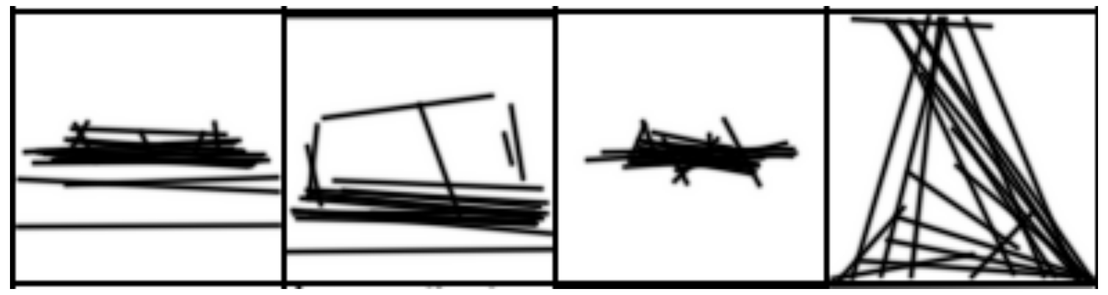
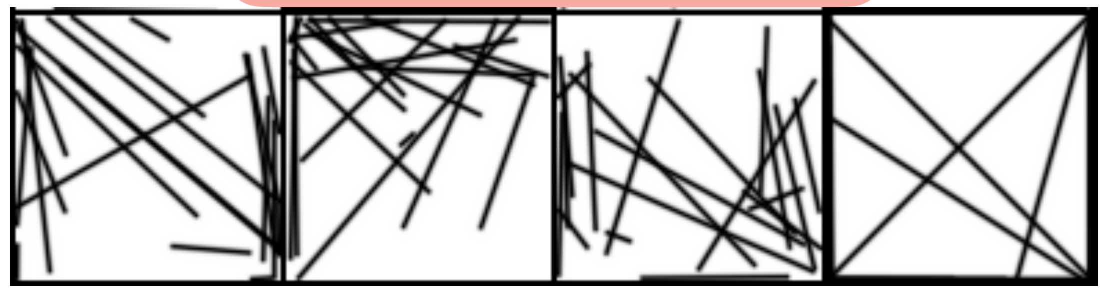
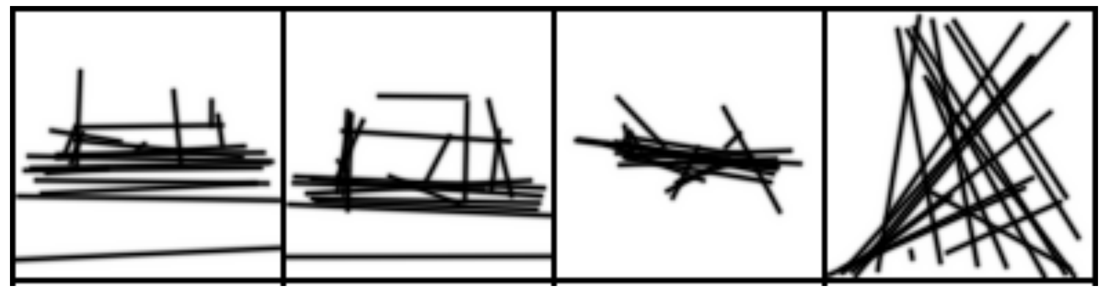
Can our agents communicate between themselves?

	$\mathrm{l}_{\mathrm{game}}$	$\mathrm{l}_{\mathrm{game}} + \mathrm{l}_{\mathrm{perceptual}}$
Original game 	71.8% (±6.1) 	69.57% (±2.6) 
OO-game same 	95.46% (±0.6) 	96.04% (±0.5) 
OO-game different 	82.72% (±0.8) 	81.09% (±0.6) 

STL-10 images, 20 lines per sketch

Experiments

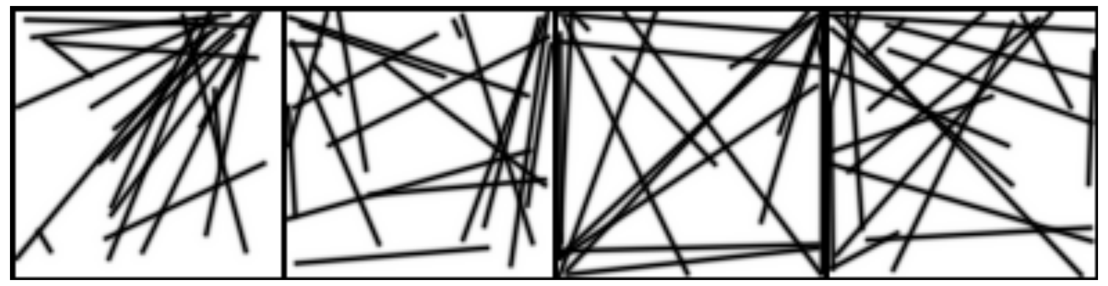
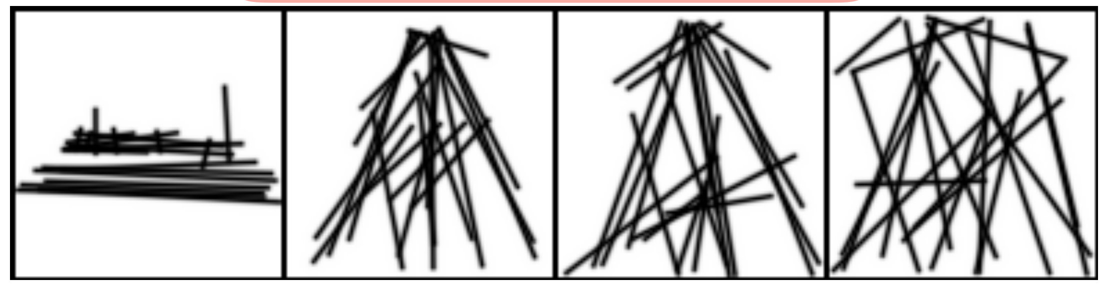
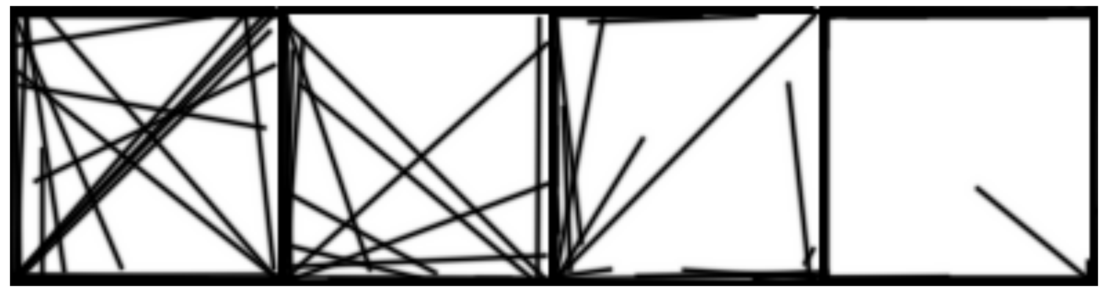
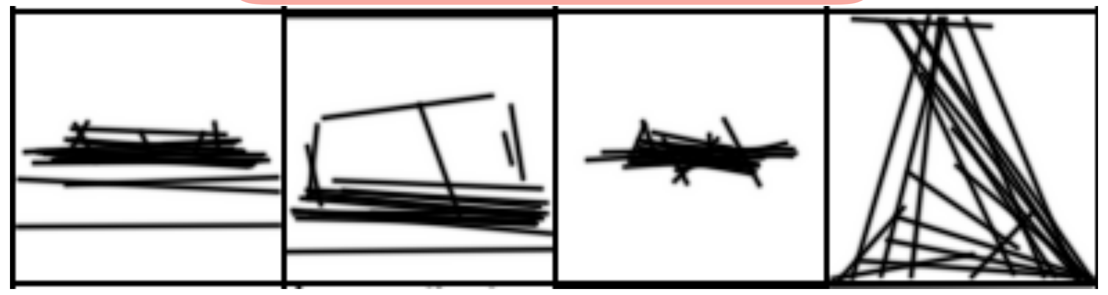
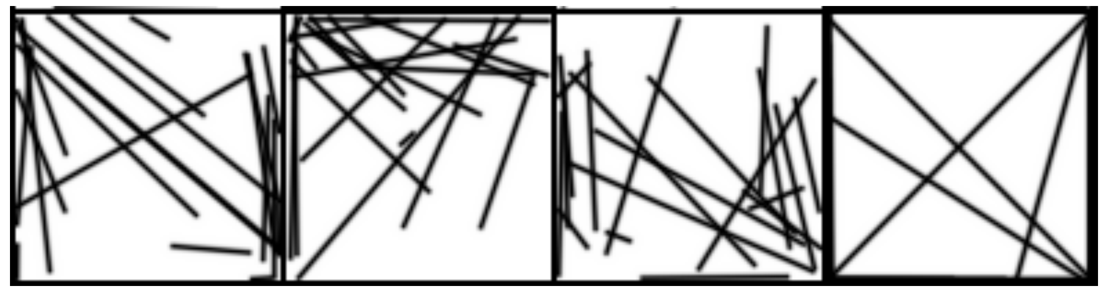
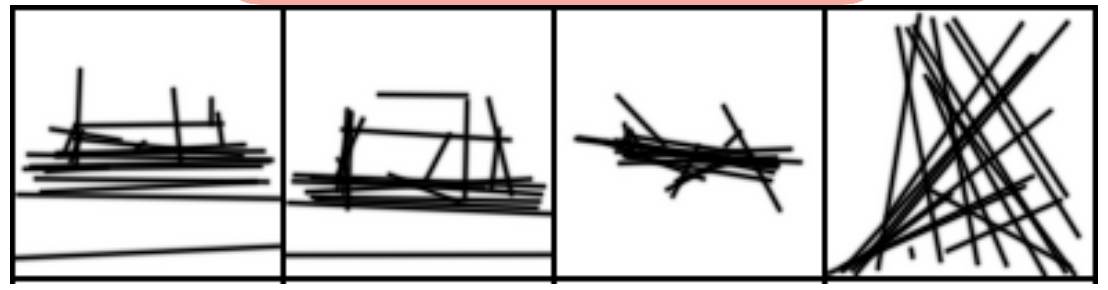
Can our agents communicate between themselves?

	l_{game}	$l_{game} + l_{perceptual}$
Original game 	71.8% (±6.1) 	69.57% (±2.6) 
OO-game same 	95.46% (±0.6) 	96.04% (±0.5) 
OO-game different 	82.72% (±0.8) 	81.09% (±0.6) 

STL-10 images, 20 lines per sketch

Experiments

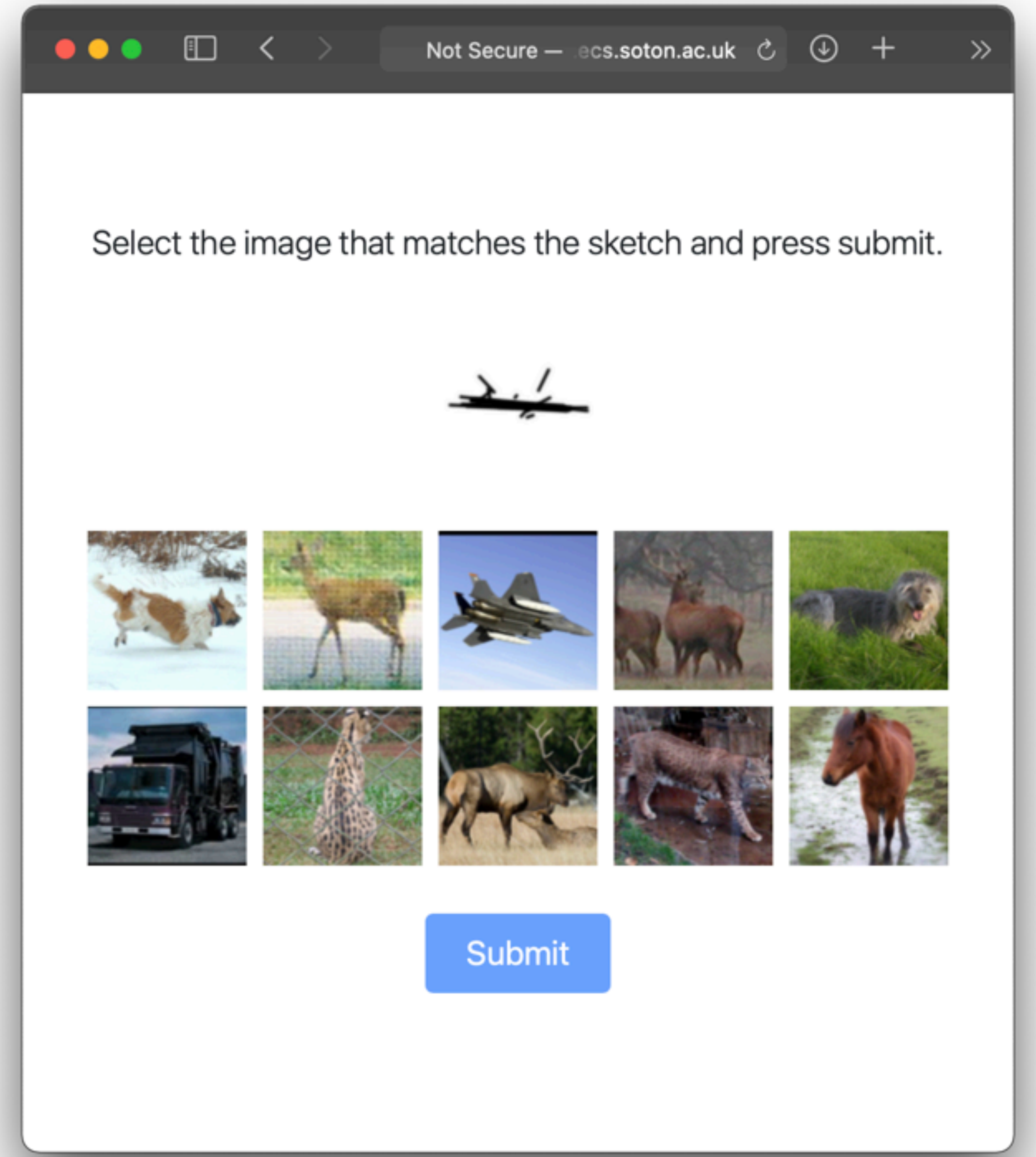
Can our agents communicate between themselves?

	l_{game}	$l_{game} + l_{perceptual}$
Original game 	71.8% (±6.1) 	69.57% (±2.6) 
OO-game same 	95.46% (±0.6) 	96.04% (±0.5) 
OO-game different 	82.72% (±0.8) 	81.09% (±0.6) 

STL-10 images, 20 lines per sketch

Experiments

Can our sender agent communicate with a Human receiver?

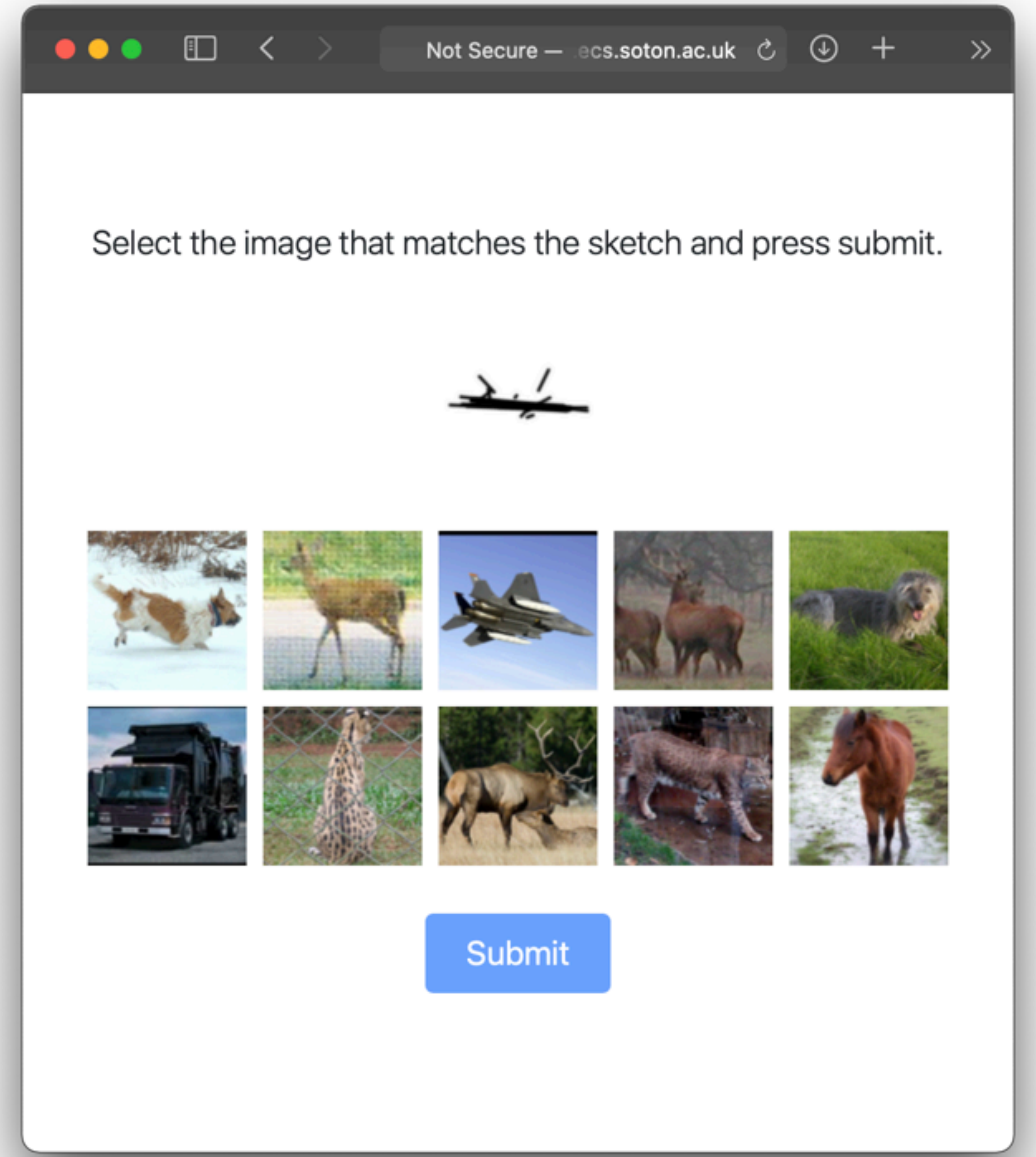


Human participants played 30 games in 5 different settings.

In total we recorded 1800 games.

Experiments

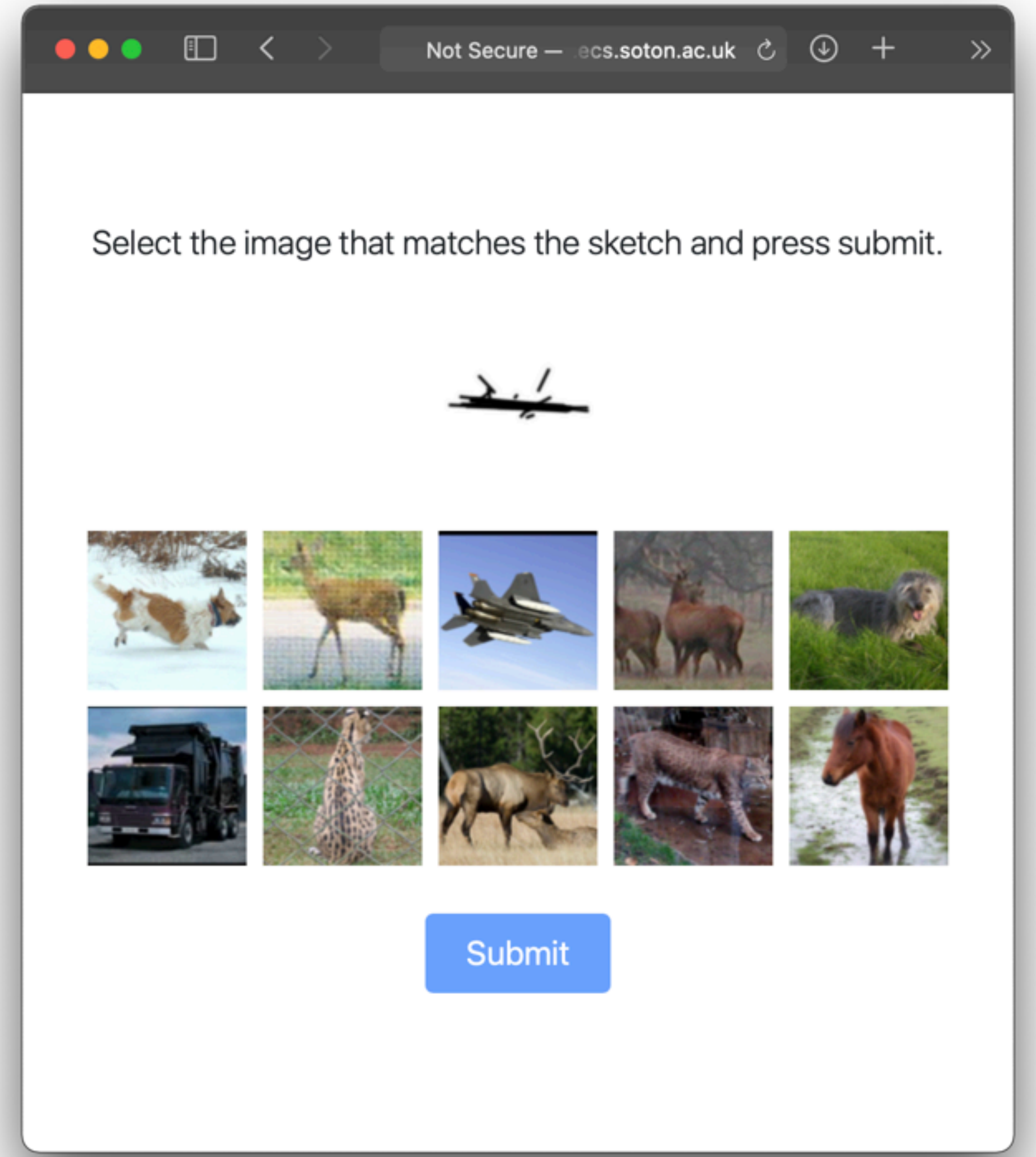
Can our sender agent communicate with a Human receiver?



Game	Loss	Lines	Agent comm. rate	Human comm. rate	Human class comm. rate
original	$l = l_{game}$	20	100%	8.3% (±5.4)	15.0% (±2.5)
original	$l = l_{game} + l_{perceptual}$	20	93.3%	38.3% (±2.5)	55.6% (±7.1)
original	$l = l_{game} + l_{perceptual}$	50	100%	37.2% (±5.9)	47.8% (±7.4)
oo diff	$l = l_{game} + l_{perceptual}$	20	83.3%	23.9% (±6.2)	23.9% (±6.2)
oo diff	$l = l_{game} + l_{perceptual}$	50	90.0%	38.9% (±9.9)	38.9% (±9.9)

Experiments

Can our sender agent communicate with a Human receiver?

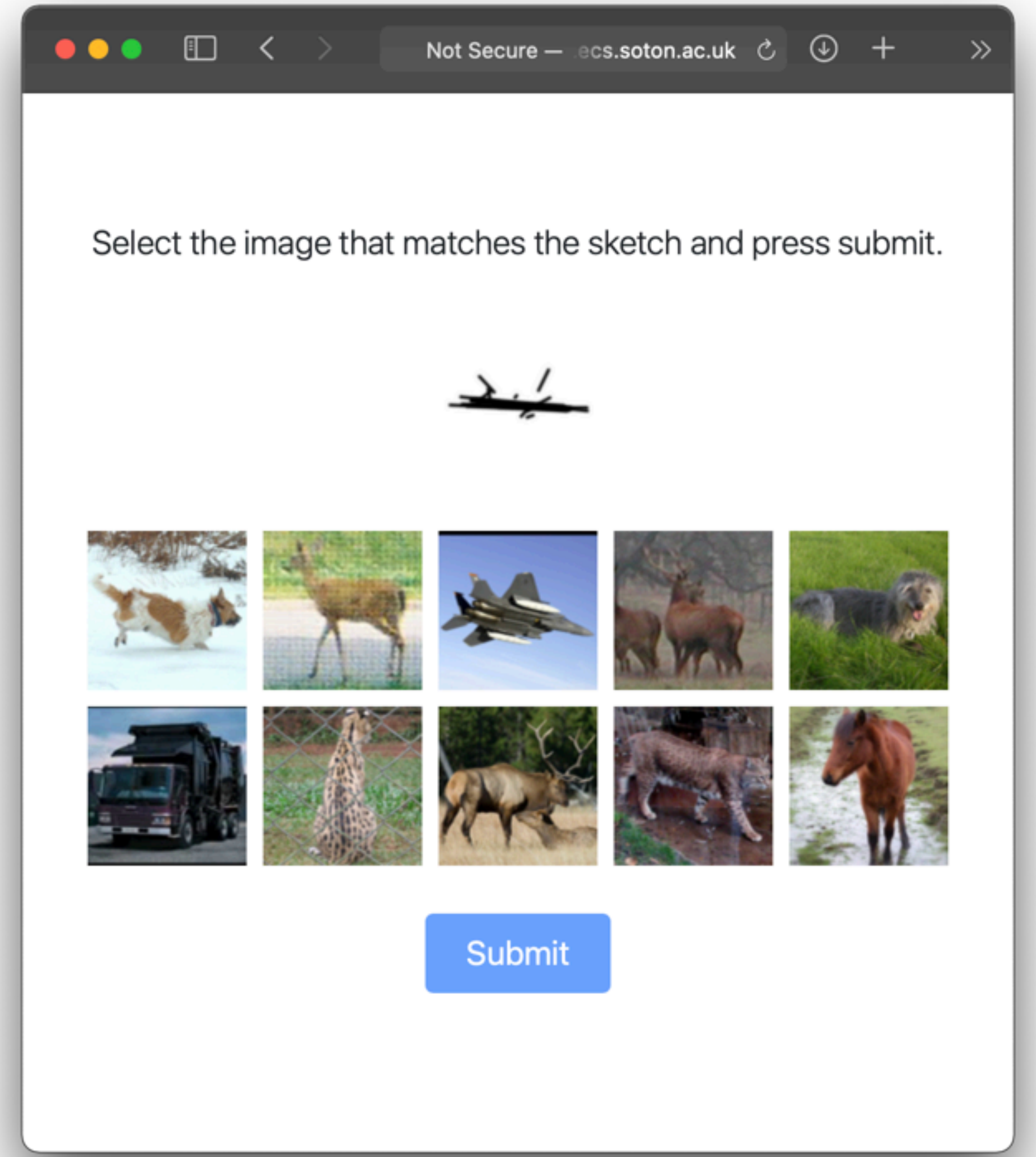


Game	Loss	Lines	Agent comm. rate	Human comm. rate	Human class comm. rate
original	$l = l_{game}$	20	100%	8.3% (±5.4)	15.0% (±2.5)
original	$l = l_{game} + l_{perceptual}$	20	93.3%	38.3% (±2.5)	55.6% (±7.1)
original	$l = l_{game} + l_{perceptual}$	50	100%	37.2% (±5.9)	47.8% (±7.4)
oo diff	$l = l_{game} + l_{perceptual}$	20	83.3%	23.9% (±6.2)	23.9% (±6.2)
oo diff	$l = l_{game} + l_{perceptual}$	50	90.0%	38.9% (±9.9)	38.9% (±9.9)

Use of the perceptual loss significantly improves the ability of a human to play the game successfully.

Experiments

Can our sender agent communicate with a Human receiver?

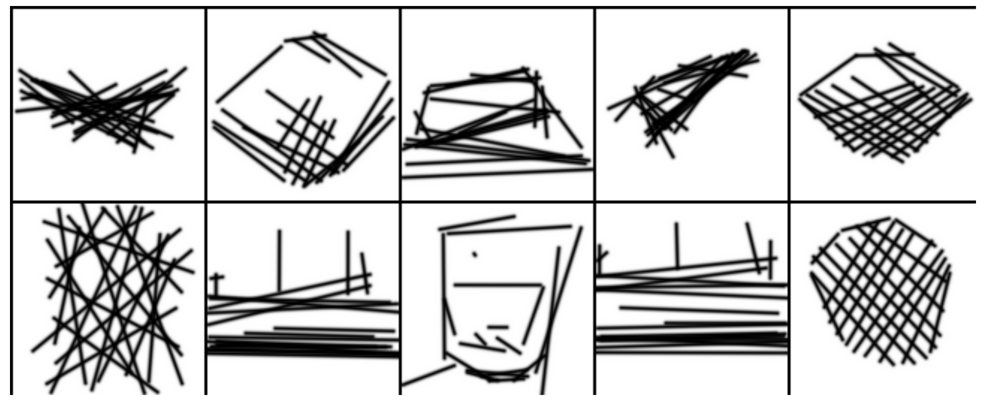


Game	Loss	Lines	Agent comm. rate	Human comm. rate	Human class comm. rate
original	$l = l_{game}$	20	100%	8.3% (±5.4)	15.0% (±2.5)
original	$l = l_{game} + l_{perceptual}$	20	93.3%	38.3% (±2.5)	55.6% (±7.1)
original	$l = l_{game} + l_{perceptual}$	50	100%	37.2% (±5.9)	47.8% (±7.4)
oo diff	$l = l_{game} + l_{perceptual}$	20	83.3%	23.9% (±6.2)	23.9% (±6.2)
oo diff	$l = l_{game} + l_{perceptual}$	50	90.0%	38.9% (±9.9)	38.9% (±9.9)

Humans are better at determining the class of the object in the sketch than recognising the specific image which matches.

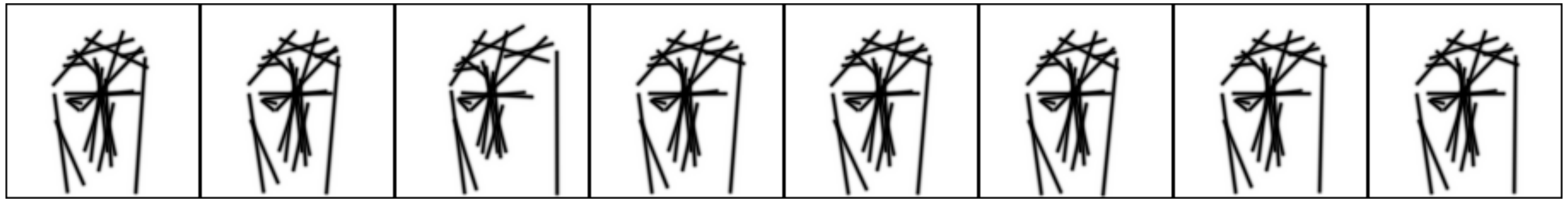
Experiments

How does a shape-bias change the sketches?

	ImageNet weights	Stylized-ImageNet weights	
	78.46% (±2.0) 	77.09% (±1.9) 	Caltech101

CelebA

ImageNet weights (99.4%)



Stylized weights (99.6%)

Experiments

Other experiments

- In the paper we also ask:
    - How does model capacity influence the communication channel?
    - Does the object-oriented setup make sketches more recognisable as the type of object?
    - How does weighting the perceptual loss change the sketches?
    - Do the models learn to pick out salient features?

Summary and Conclusions

Summary and Conclusions

- We have demonstrated that:
  - It is possible to build agents that **successfully learn to communicate through sketches**.
  - We can train the agents through **self-play** using **end-to-end gradient-based optimisation**.

Summary and Conclusions

- We have demonstrated that:
  - It is possible to build agents that **successfully learn to communicate through sketches**.
  - We can train the agents through **self-play** using **end-to-end gradient-based optimisation**.
  - Appropriate **inductive biases** can be added during training which encourage the agents to communicate in a **visibly more interpretable manner**.
  - Further, through a **study with human participants** we have demonstrated that it is possible for a trained sketching agent to **successfully communicate with humans**.

UNIVERSITY OF Southampton

Summary and Conclusions

What next?

- Improved drawing (curves, shapes, etc.).
- Improved models: Could a more advanced visual system be incorporated?
- Improved understanding: explore what groups of strokes “mean”, explore if the sketches produced could be considered to be “compositional”.

Thank you for listening!