# Equivariant Transformer Networks

(Poster 18)

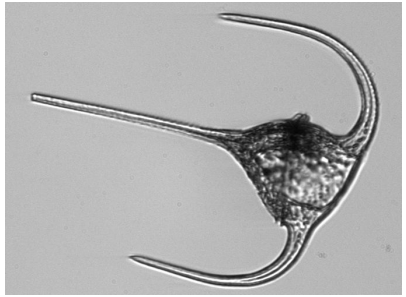
**Kai Sheng Tai**, Peter Bailis & Gregory Valiant
Stanford University

github.com/stanford-futuredata/**equivariant-transformers**

# Goal: Transformation-invariant models

- How can we learn models that are invariant to certain input transformations?
- Relevant to many application domains:

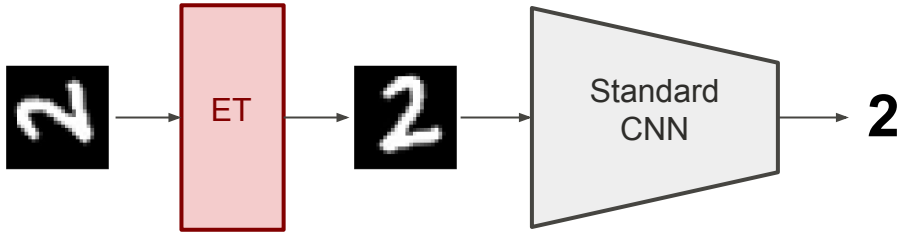
astronomical objects


plankton micrographs

traffic signs

- In this work, we explore alternatives to data augmentation
- How can we build invariances directly into network architectures?
  [Group Equivariant CNNs (Cohen+'16, Dieleman+'16), Harmonic Networks (Worrall+'17), etc.]
- Can we achieve invariance while reusing off-the-shelf architectures?
  [Spatial Transformer Networks (Jaderberg+'15)]

# Equivariant Transformer Layers



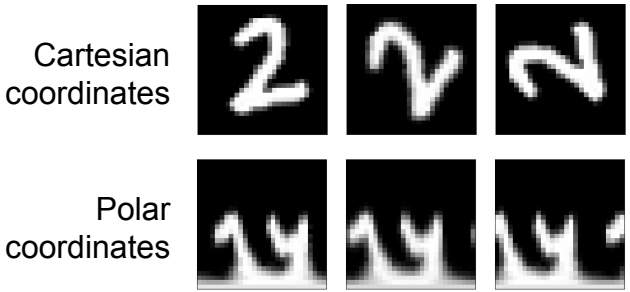

- An Equivariant Transformer (ET) is a differentiable image-to-image mapping
- **Key property** (“**local invariance**”):
  - **all** transformed versions of a base image are mapped to the **same** output image
- **Requirement**:
  - family of transformations forms a Lie group:
    transformations are invertible, differentiable wrt a real-valued parameter
  - includes many common families of transformations:
    translation, rotation, scaling, shear, perspective, etc.

# Key ideas

1. Standard convolutional layers are **translation-equivariant**
    - i.e., input translated by $\theta \rightarrow$ output translated by $\theta$

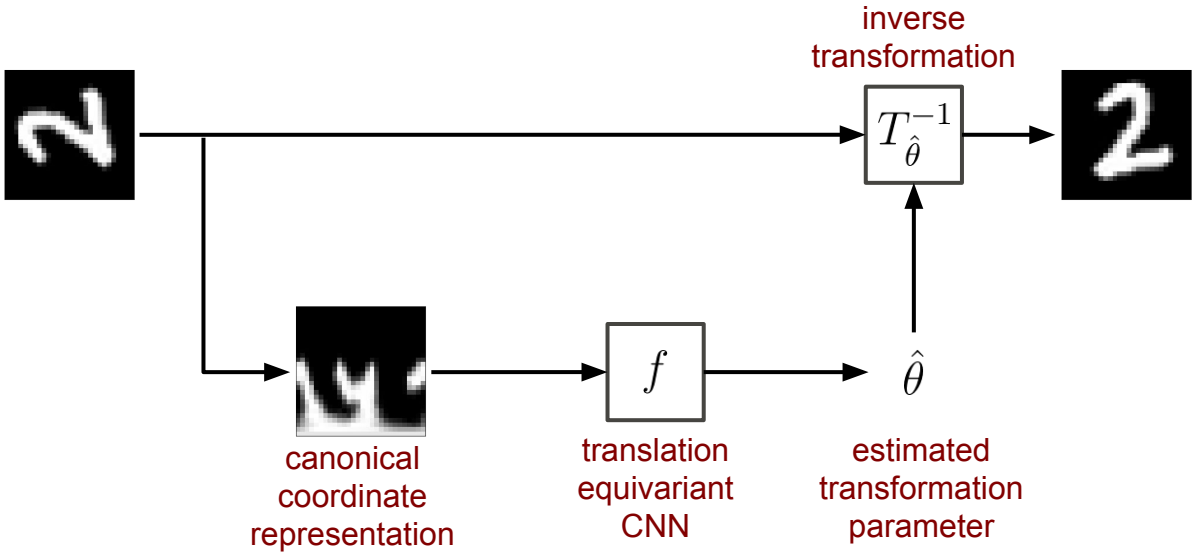
# Key ideas

1. Standard convolutional layers are **translation-equivariant**
   - i.e., input translated by $\theta \rightarrow$ output translated by $\theta$
2. Specialized coordinates turn smooth transformations into **translation**
   - Example (rotation): in polar coordinates, rotation appears as translation by angle $\theta$




   - This can be generalized to other smooth transformations using *canonical coordinate systems for Lie groups* (Rubinstein+'91)

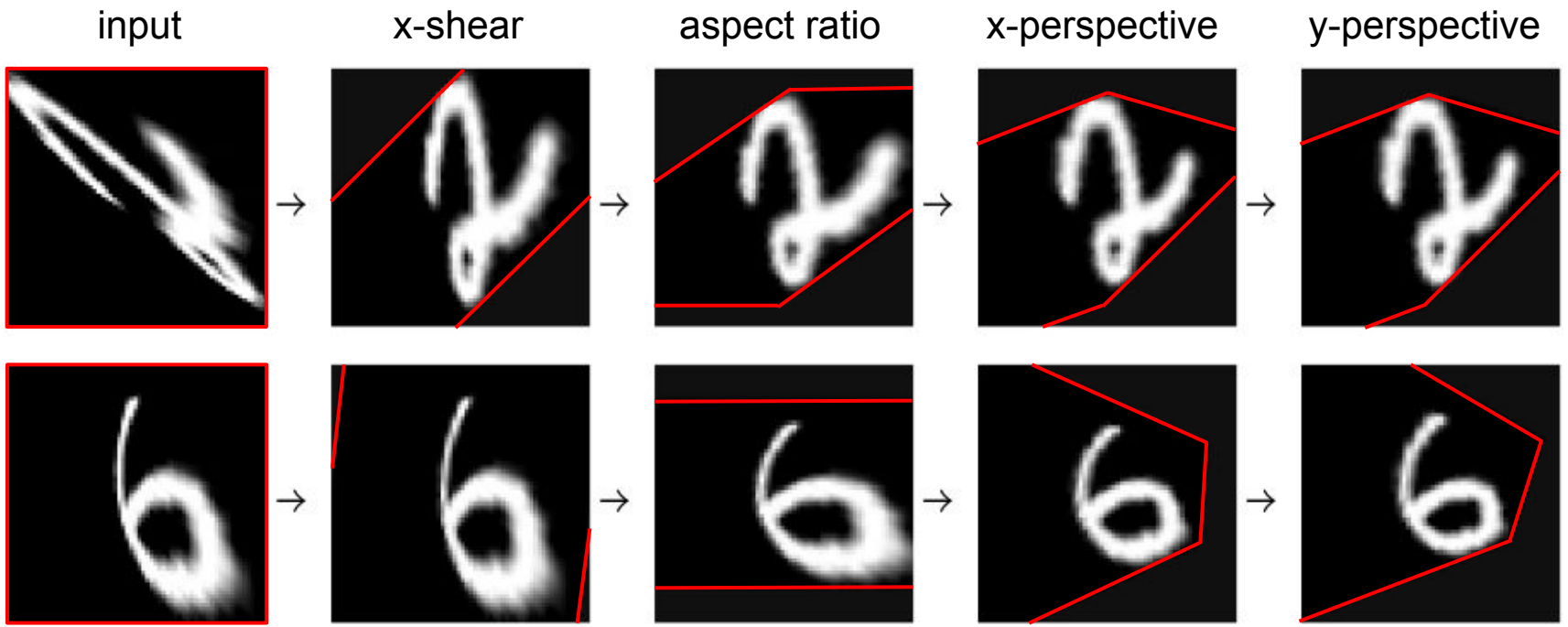
# ETs are locally invariant by construction



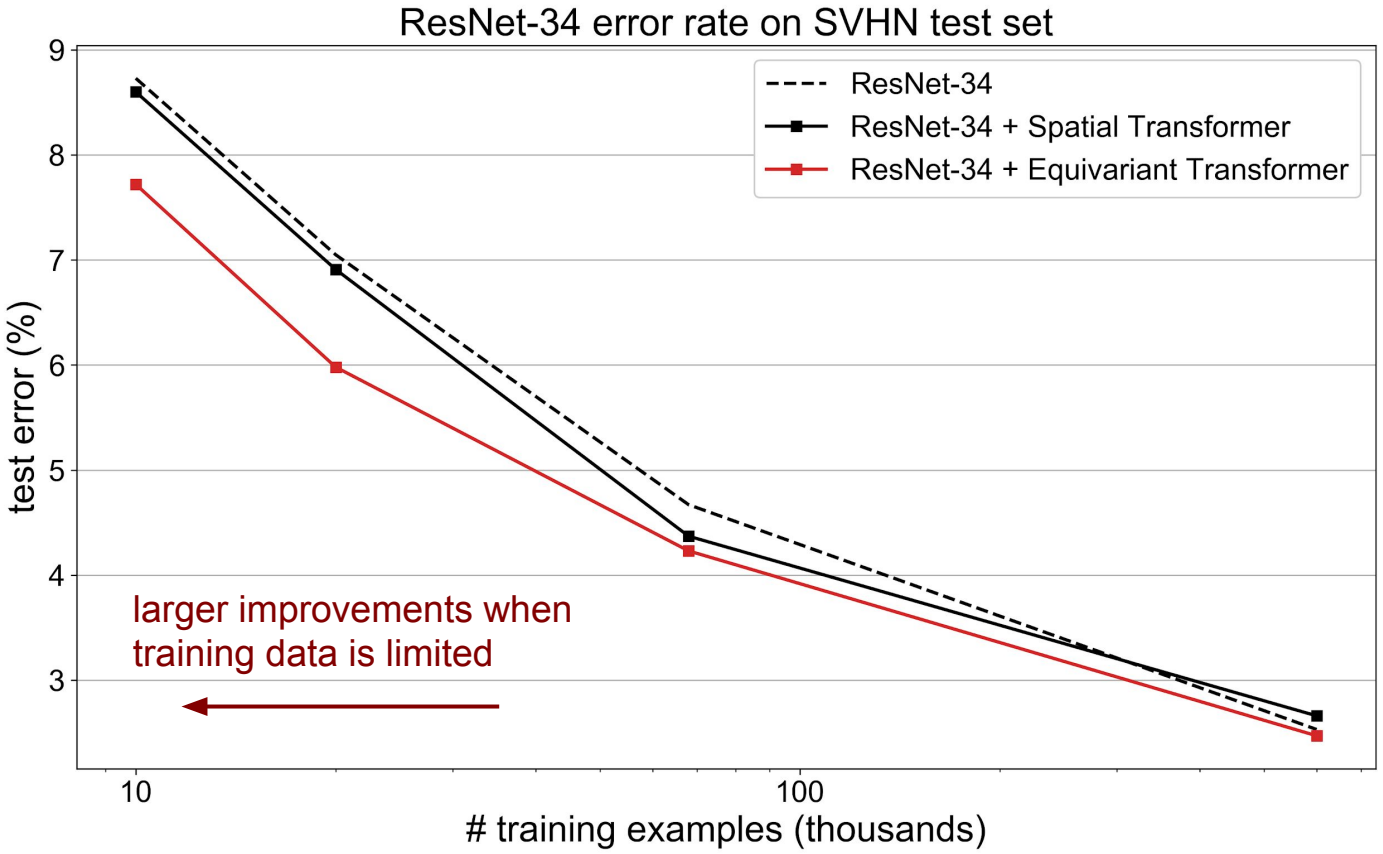

- Equivariance **guarantees** that an additional transformation of $\theta$ causes the estimated parameter to be increased by $\theta$
- The output is therefore invariant to transformations of the input
- We implement transformation with differentiable grid resampling (Jaderberg+'15)

# Compositions of ETs handle more complicated transformations

- Since ETs map images to images, they can be composed sequentially



# ETs improve generalization



# Takeaways

- Equivariant Transformers build transformation invariance into neural network architectures
- **Main ideas:**
  - Canonical coordinates let us tailor ET layers to specific transformation groups
  - Image-to-image interface lets us compose ETs to handle more complicated transformation groups

Poster #18
kst@cs.stanford.edu

## Try it yourself!

github.com/stanford-futuredata/**equivariant-transformers**