Geometry-Aware Neural Rendering

Josh Tobin, OpenAI Robotics, Pieter Abbeel
Robots need to understand the world to act
Visuomotor policies

Noisy observations

Action

\[ \pi(o) \mapsto a \]
State estimation + control

Noisy observations → State representation $S$ → Policy $\pi(s) \mapsto a$ → Action
How to model more complex states?

From…

…To
Model the state of the world implicitly

The Neural Rendering problem

Context viewpoints

\( v^1 \)

\( v^2 \)

Arbitrary query viewpoint

\( v^q \)
Motivation: Generative Query Nets (GQN)

Motivation: Generative Query Nets (GQN)

Motivation: Generative Query Nets (GQN)

Our work: extends GQN to…

- Higher-dimensional images
- Objects with many degrees of freedom
- Wider range of more realistic objects
Our approach: add attention

\[ v^1 \rightarrow r^1 \rightarrow r \rightarrow h_1 \rightarrow h_2 \rightarrow \cdots \rightarrow h_L \]

\[ v^2 \rightarrow r^2 \rightarrow r \rightarrow h_1 \rightarrow h_2 \rightarrow \cdots \rightarrow h_L \]

Representation Network

Generation Network
Our approach: add attention

Representation Network

Generation Network

\[ v^q \]

\[ h_1 \rightarrow h_2 \rightarrow \cdots \rightarrow h_L \]

\[ z \]
Our approach: add attention

Representation Network

Generation Network

$\nu^1$ $r^1$ $h_1$ $h_2$ $\cdots$ $h_L$ $\nu^q$ $z$
Our approach: add attention

Representation Network

Attention Mechanism

Generation Network
Our approach: add attention
Attention mechanism: intuition

Context image

Target image
Attention mechanism: intuition

Context image

Target image
Attention mechanism: intuition

Context image

Target image
Attention mechanism: intuition

Context image

Target image
Attention mechanism: intuition

Context image

Target image
Attention mechanism: intuition

Context image

Target image
Attention mechanism: intuition

Context image

Target image
Attention mechanism: intuition

$O(n^2)$ comparisons for each pixel
Epipolar geometry
Epipolar geometry

Context image

Target image

True 3D point
Epipolar geometry

Context image

Target image

Image of 3D point

$O(n)$ comparisons for each pixel
Epipolar attention

Representation Network

Attention Mechanism

Generation Network

$v^1$

$r^1$

$v^q$

$h_1 \rightarrow h_2 \rightarrow \cdots \rightarrow h_L$

$z$
Epipolar attention
Epipolar attention

Attention mechanism

Generation Network

$v^1$

$v^q$

$h_1$

$v^1$

$v^q$

$h_1$
Epipolar attention

Generation Network

Attention mechanism

$v_1 \rightarrow r_1 \rightarrow v_1 \rightarrow v_q \rightarrow h_1$
Epipolar attention
Epipolar attention

Generation Network
Epipolar attention

[Diagram showing epipolar attention mechanism]

$v^1 \rightarrow r^1 \rightarrow h_1$,

$v^q \rightarrow$ Attention mechanism

Generation Network
Epipolar attention

Generation Network

Attention mechanism

$v^1 \rightarrow r^1 \rightarrow e^1 \rightarrow h_1$
Epipolar attention

Attention mechanism

\[ v^q \]

Generation Network

\[ e^1 \]

\[ h_1 \rightarrow \cdots \rightarrow h_{l-1} \]
Epipolar attention

Generation Network

Attention mechanism

$v^1$, $r^1$, $e^1$, $h_l$, $h_{l-1}$
Epipolar attention

\[ \nu^1 \xrightarrow{} r^1 \rightarrow e^1 \rightarrow h_1 \rightarrow \cdots \]

Attention mechanism

Generation Network
Epipolar attention

\[ v^1 \rightarrow r^1 \rightarrow e^1 \rightarrow h_{l-1} \rightarrow a^1_l \rightarrow h_1 \rightarrow \cdots \]

Attention mechanism

Generation Network
Epipolar attention

\[ v^1 \to r^1 \to e^1 \to a^1_l \to h_1 \to \ldots \to h_{l-1} \]

\[ \nu^q \]

Attention mechanism

Generation Network
Epipolar attention

<table>
<thead>
<tr>
<th>Generation Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v^1$</td>
</tr>
<tr>
<td>$r^1$</td>
</tr>
<tr>
<td>$e^1$</td>
</tr>
<tr>
<td>$a^1_l$</td>
</tr>
<tr>
<td>$h_1$</td>
</tr>
<tr>
<td>$h_{l-1}$</td>
</tr>
<tr>
<td>$h_l$</td>
</tr>
</tbody>
</table>

Attention mechanism
Examples

Context | Target | GQN | E-GQN
--- | --- | --- | ---

[Images of examples]
Examples

Context

Target

GQN

E-GQN
Examples
Conclusion

• Geometrically-inspired neural network primitives improve implicit 3D understanding
Questions?

• arXiv: 1911.04554

• Poster session: Wed Dec 11th 05:00 - 07:00 PM @ East Exhibition Hall B + C #128

• Twitter: @josh_tobin_