Differentiable Spatial Planning using Transformers

ICML 2021

Webpage: https://devendrachaplot.github.io/projects/spatial-planning-transformers
Spatial Planning

- Input
  - spatial obstacle map
  - goal location
  - starting location

- Output
  - Shortest path to the goal.

(a) Navigation

(b) Manipulation

\( \theta_1 \)

\( \theta_2 \)
Why learn to plan?
Why learn to plan?

Exploit statistical regularity in data
Why learn to plan?

Exploit statistical regularity in data

Tackle unknown maps

Map images from the Gibson dataset (Xia et al. CVPR 2018)
Why Transformers?
Why Transformers?

- **Prior Methods**: Inductive bias on local value propagation using CNNs
  - VIN [1]: CNNs with tied weights
  - GPPN [2]: Convolutional LSTMs
Why Transformers?

- **Prior Methods**: Inductive bias on local value propagation using CNNs
  - VIN [1]: CNNs with tied weights
  - GPPN [2]: Convolutional LSTMs

- **Long distance value propagation**
  - Value can be propagated between two distant points with no obstacle between them
  - Transformers are well suited, can attend to arbitrary cells

---

[1] Value Iteration Networks (VIN) [Tamar et al. NeurIPS 2016]
[2] Gated Path-Planning Networks (GPPN) [Lee et al. ICML 2018]
Planning with known maps
Planning with known maps

Navigation

Goal ($g$)  Map ($m$)  Planning model  Predicted Action Distances ($\hat{y}$)
Planning with known maps

**Navigation**
- Goal ($g$)
- Map ($\hat{m}$)

**Planning model**
- Spatial Planning ($f_P$)
- Predicted Action Distances ($\hat{y}$)

**Manipulation**
- Convert to configuration space
- Goal ($g$)
- Map ($\hat{m}$)

**Planning model**
- Spatial Planning ($f_P$)
- Predicted Action Distances ($\hat{y}$)

**Navigation** and **Manipulation** both involve converting the goal and map information into a planning model, which then predicts the action distances.

**Equations**:
- $m$ (map)
- $g$ (goal)
- $f_P$ (spatial planning function)
- $\hat{y}$ (predicted action distances)
Spatial Planning Transformer (SPT)
Spatial Planning Transformer (SPT)
Spatial Planning Transformer (SPT)

Input map and goal ($x$)

Encoder $E$

$M = \text{Map Size}$

2-layer Conv (1x1)

$M = \text{Map Size}$
Spatial Planning Transformer (SPT)

M = Map Size

Input map and goal (x)

2-layer Conv (1x1)

Encoder E

$\mathcal{M}^2$

Transformer $T$

Input Encoding ($x_I$)
Spatial Planning Transformer (SPT)

- **Input Encoding** ($x_I$)
- **Output Encoding** ($x_O$)
- **Positional Encoding** ($p$)
- **Transformer Layer** ($f_{TL}$)
- **Encoder** ($E$)
- **Transformer** ($T$)
- **Transformer Layers** ($N = 5$)
- **Map Size** ($M$)

Symbols:
- $x_I$: Input Encoding
- $x_O$: Output Encoding
- $p$: Positional Encoding
- $E$: Encoder
- $T$: Transformer
- $N$: Number of Transformer Layers
- $M$: Map Size

Diagram elements:
- 2-layer Conv (1x1)
- Flatten
- Transformer Layer
- Input map and goal ($\lambda$)

Mathematical expressions:
- $M = \text{Map Size}$
Spatial Planning Transformer (SPT)

$M = \text{Map Size}$

Input map and goal ($x$)

Predicted distance ($\hat{y}$)

2-layer Conv (1x1)

Encoder $E$

Flatten

$M = \text{Map Size}$

Decoder $D$

Transformer $T$

$N = 5$ Transformer Layers

Positional Encoding ($p$)

Input Encoding ($x_I$)

Output Encoding ($x_O$)

Transformer Layer ($f_{TL}$)

$Z_1 + Z_2 + \ldots + Z_M$

$0.1 \ | \ 0.9 \ | \ \ldots \ | \ 0.2$

$M^2$

Spatial Planning Transformer (SPT)
Spatial Planning Transformer (SPT)
Training SPT with synthetic data
Training SPT with synthetic data

Input map and goal ($x$)

SPT

Predicted distance ($\hat{y}$)

Ground truth ($y^*$)

MSE Loss
Training SPT with synthetic data

Input map and goal ($x$)

Predicted distance ($\hat{y}$)

Ground truth ($y^*$)

MSE Loss

SPT
Training SPT with synthetic data

Input map and goal ($x$) → Predicted distance ($\hat{y}$) → Ground truth ($y^*$)

MSE Loss
Training SPT with synthetic data

Input map and goal ($x$)  
Predicted distance ($\hat{y}$)  
Ground truth ($y^*$)

SPT

MSE Loss
Training SPT with synthetic data

- Input map and goal ($x$)
- Predicted distance ($\hat{y}$)
- Ground truth ($y^*$)

MSE Loss
Planning with unknown maps
Planning with unknown maps

Goal \((g)\)  
Observation Space \((o)\)  
Mapping \((f_M)\)  
Predicted Map \((\hat{m})\)  
Spatial Planning \((f_P)\)  
Predicted Action Distances \((\hat{f})\)
Planning with unknown maps

**Navigation**
- Goal \((g)\)
- Observation Space \((o)\)
- Mapping \((f_M)\)
- Predicted Map \((\hat{m})\)
- Spatial Planning \((f_P)\)
- Predicted Action Distances \((\hat{y})\)

**Manipulation**
- Goal \((g)\)
- Observation Space \((o)\)
- Mapping \((f_M)\)
- Predicted Map \((\hat{m})\)
- Spatial Planning \((f_P)\)
- Predicted Action Distances \((\hat{y})\)
Experiments
Experiments

- Baselines:
  - Value Iteration Networks (VIN) [Tamar et al. NeurIPS 2016]
  - Gated Path-Planning Networks (GPPN) [Lee et al. ICML 2018]
Experiments

• Baselines:
  • Value Iteration Networks (VIN) [Tamar et al. NeurIPS 2016]
  • Gated Path-Planning Networks (GPPN) [Lee et al. ICML 2018]
• Metric: Planning accuracy
Experiments

- Baselines:
  - Value Iteration Networks (VIN) [Tamar et al. NeurIPS 2016]
  - Gated Path-Planning Networks (GPPN) [Lee et al. ICML 2018]
- Metric: Planning accuracy
- Datasets
Experiments

• Baselines:
  • Value Iteration Networks (VIN) [Tamar et al. NeurIPS 2016]
  • Gated Path-Planning Networks (GPPN) [Lee et al. ICML 2018]
• Metric: Planning accuracy
• Datasets
  • In-distribution (0-5 obstacles)
Experiments

- Baselines:
  - Value Iteration Networks (VIN) [Tamar et al. NeurIPS 2016]
  - Gated Path-Planning Networks (GPPN) [Lee et al. ICML 2018]
- Metric: Planning accuracy
- Datasets
  - In-distribution (0-5 obstacles)
  - Out-of-distribution
    - More obstacles (15-20 obstacles)
    - Real-world maps (Gibson dataset) [Xia et al. CVPR 2018]
Results

In-distribution

Planning Accuracy

VIN  80.0
GPPN 92.3
SPT  99.4

Ground Truth Input Map and Goal

In-distribution
Results

Out-of-distribution

<table>
<thead>
<tr>
<th>Method</th>
<th>Planning Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIN</td>
<td>60.6</td>
</tr>
<tr>
<td>GPPN</td>
<td>87.0</td>
</tr>
<tr>
<td>SPT</td>
<td>95.2</td>
</tr>
</tbody>
</table>

Ground Truth

Input Map and Goal

Out-of-distribution
Results

Unknown maps

<table>
<thead>
<tr>
<th>Method</th>
<th>Planning Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIN</td>
<td>57.1</td>
</tr>
<tr>
<td>GPPN</td>
<td>63.9</td>
</tr>
<tr>
<td>SPT</td>
<td>82.3</td>
</tr>
</tbody>
</table>

Ground Truth and Goal

Input Map

Predicted Map

Ground Truth Map

Predicted Distance

Ground Truth Distance
Differentiable Spatial Planning using Transformers
Devendra Singh Chaplot, Deepak Pathak, Jitendra Malik
ICML 2021

Webpage: https://devendrachaplot.github.io/projects/spatial-planning-transformers

Thank you

Devendra Singh Chaplot
Webpage: http://devendrachaplot.github.io/
Email: chaplot@cs.cmu.edu
Twitter: @dchaplot