1 Video Information Propagation

**Aim:** To devise a general neural network approach for structured information propagation across video frames.

**Useful for many video applications:** Video object segmentation, Semantic video segmentation, Video color propagation etc.

**Challenges:**
- To account for scene and camera motion.
- To process large number of video frames.

**Approach:** A neural network with video adaptive receptive fields and some tricks for computational efficiency.

**Code:** [varunjampani.github.io/vpn](http://varunjampani.github.io/vpn)

**VPN** is the first neural network approach for video information propagation, can propagate any type of video content, handle large number of input frames and have favorable runtime.

2 Video Propagation Networks with Learnable Bilateral Filters

**Bilateral Convolution Layer (BCL) [1, 2] for forward information propagation:**
1. **Split** previous frame results onto 6 dimensional permutohedral lattice constructed with \((x,y,\gamma,\alpha,\nu)\) features.
2. **Convolve** the high-dimensional signal with learnable filters [1].
3. **Slice** the resulting signal onto the present frame.

**Advantages of VPN:**
- General applicability
- Online propagation
- End-to-end trainable
- Long range and image-adaptive filtering
- Runtime

**Comparison to spatial convolutions:**
- Efficient sparse high-dimensional filtering with hash table.
- Image adaptive receptive fields.
- Input and output points can be different.

**Some tricks for efficient video propagation:**
- **Split** only a random set of points or superpixels from previous frames.
- **Use** 1x1 convolutions in high-dimensional space as building neighborhood.

**Fragment-based algorithms:**
- **Random Set of Points vs.**
- **Splat only a**

3 Application: Object Segmentation

**Task:** Given an object mask for the first video frame, predict the corresponding object masks for the entire video.

**State-of-the-art results on DAVIS dataset [3] at the time of submission with favorable runtime despite VPN not using any strong deep appearance model.**

VPN can be easily integrated into other deep architectures. We show a boost in performance when augmenting VPN with DeepLab [4] model (VPN-DeepLab).

VPN results on DAVIS [3] Dataset

<table>
<thead>
<tr>
<th>PSNR</th>
<th>Results on DAVIS [3] Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.15</td>
<td>DAVIS [3]</td>
</tr>
<tr>
<td>26.15</td>
<td>DAVIS [3]</td>
</tr>
<tr>
<td>26.15</td>
<td>DAVIS [3]</td>
</tr>
<tr>
<td>26.15</td>
<td>DAVIS [3]</td>
</tr>
</tbody>
</table>

4 Application: Semantic Video Segmentation

**Task:** Refine a semantic video segmentation result obtained by a CNN run on each frame independently.

**VPN works better yet faster compared to video CRF techniques such as PSO-CRF [6].**

Using optical flow to normalize the bilateral features (VPN-Flow) results in IOU improvements.

5 Application: Video Color Propagation

**Task:** Propagate color from first frame to the remaining video frames in a given grayscale video.

**VPN works faster and better compared to traditional optimization based techniques [7].**

References: