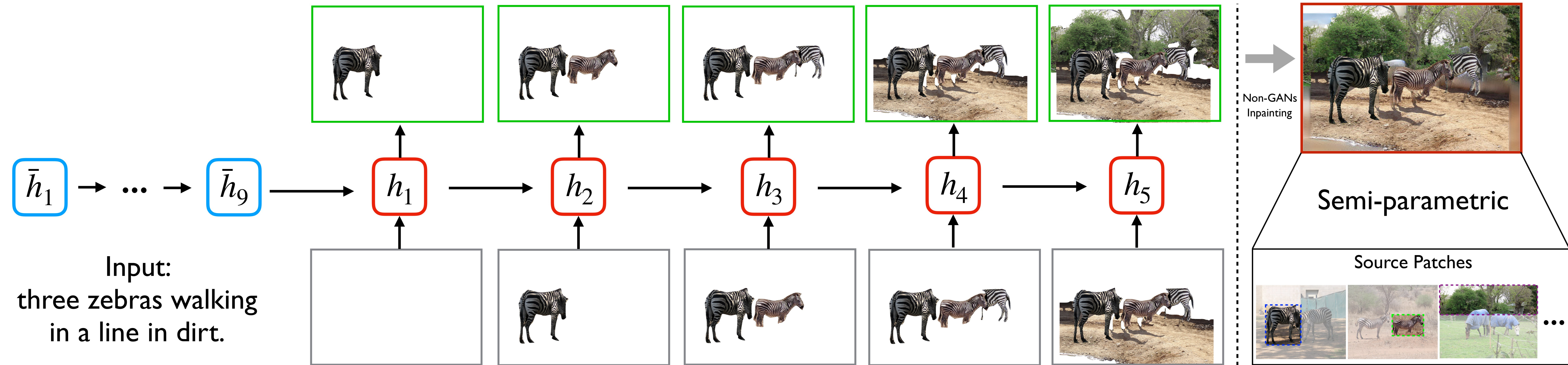


## Contributions

- Text2Scene, a Non-GANs approach for image generation from text, providing more interpretable results;
- A unified sequence-to-sequence framework supporting the generation of different forms of scene representations, including cartoon-like scenes, object layouts corresponding to real images, and synthetic images;
- Comparable with state-of-the-art GAN-based methods on automatic metrics and superior based on human judgments.

## Text2Scene: A Sequence-to-Sequence Framework



## Goal: Generating Scenes from Text

Input	Text2Scene	Ground-truth
Mike is surprised at the duck. The duck is standing on the grill. Jenny is running towards Mike and the duck.		
A guy on a motorcycle with some people watching.		
Several elephants walking together in a line near water.		

## I: Abstract Scene Generation

	Zitnick et al.	Text2Scene	Ground-truth
Human Evaluation	0.555	0.644	0.919

Input	Zitnick et al.	Text2Scene	Ground-truth
Mike went down the slide fast. Jenny is worried that Mike is hurt. Jenny is wearing a chef hat.			
Mike is angry at Jenny. Jenny is sad that Mike took the frisbee. The pizza is on the table.			
Jenny is holding a bucket and shovel. Mike fell off the swingset. There is rain and lightning in the sky			

## II: Layout Generation on COCO

Input	Text2Scene	Ground-truth Layout	Ground-truth Image
A happy couple is cutting a decorated cake.			
Two men and one woman in front of an elephant.			
Three people riding on the backs of elephants.			
A laptop computer a keyboard and two monitors.			

## III: Image Generation on COCO

	Text2Scene > SG2IM (CVPR 2018)	Text2Scene > HDGAN (CVPR2018)	Text2Scene > AttnGAN (CVPR 2018)
Human Evaluation	0.767	0.869	0.759

Input	SG2IM	AttnGAN	Text2Scene	Ground-truth
A woman sitting on a bench with an umbrella on her head.				
A car bridge going over a commuter train.				
A room with a TV and some different types of couches.				