

# Supplementary Materials for CookGAN: Causality based Text-to-Image Synthesis

Bin Zhu  
City University of Hong Kong  
binzhu4-c@my.cityu.edu.hk

Chong-Wah Ngo  
City University of Hong Kong  
cscwno@cityu.edu.hk

## 1. Additional Generation Examples

### 1.1. Content Manipulability





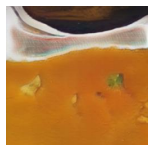











Recipe Manipulation	Concrete Operation	GT	Before Manipulation	After Manipulation
Add Ingredients	+ Carrot			
				
Minus Ingredients	- Carrot			
				
Replace Ingredients	Carrot → lettuce			
				

Figure 1. Examples of generated images by CookGAN with different recipe manipulation operators.

## 1.2. Visual Quality

Recipe	GT	CookGAN	StepGAN	IngredientGAN	StackGAN++
<p><b>Deli-Style Smoked Turkey Panini</b></p> <p>Italian bread; Turkey breast; tomato; Dijon Mustard; miracle whip...</p> <ol style="list-style-type: none"> <li>Heat panini grill.</li> <li>Fill bread slices with VELVEETA, turkey and tomatoes to make 4 sandwiches.</li> <li>Mix dressing and mustard until...</li> </ol>					
<p><b>Chocolate Chunk-Everything Cookies</b></p> <p>flour; baking soda; butter; brown sugar; eggs; Chocolate; rice cereal; Pecans; pretzels ...</p> <ol style="list-style-type: none"> <li>Heat oven to 375F.</li> <li>Mix flour and baking soda until blended.</li> <li>Beat butter and sugars in large bowl with mixer until light and ...</li> </ol>					
<p><b>Chicken Breast with Green Peppers</b></p> <p>chicken breast; green pepper; soy sauce; Katakuriko; ginger...</p> <ol style="list-style-type: none"> <li>Diagonally chop the chicken breast into bite-sized pieces, then soak in sake and soy sauce.</li> <li>Rough chop the green peppers, saute in a frying pan, season with salt...</li> </ol>					
<p><b>Cheesy Scalloped Potatoes</b></p> <p>mayonnaise; flour; salt; salt; pepper; milk; cheese; parsley; chives; potato...</p> <ol style="list-style-type: none"> <li>Combine 1/2 cup mayonnaise and next 3 ingredients in a saucepan.</li> <li>Gradually add milk, and cook, stirring constantly, over medium-low heat 8019 minutes or until thicken ...</li> </ol>					
<p><b>SeaShell Casserole</b></p> <p>shell pasta; beef; onion; tomato sauce; mushroom soup; cheese; pepper; salt; garlic powder...</p> <ol style="list-style-type: none"> <li>Preheat oven to 350 degrees Fahrenheit.</li> <li>Spray a large casserole dish with cooking spray.</li> <li>In a large skillet, brown together ground beef and onion...</li> </ol>					
<p><b>Tropical Strawberry Smoothie</b></p> <p>strawberries; coconut; pineapple juice; banana; ice...</p> <ol style="list-style-type: none"> <li>Add all ingredients into a blender.</li> <li>Blend on high speed until ice is crushed and drink is smooth.</li> </ol>					

Figure 2. Comparison of food images generated by CookGAN, StepGAN, IngredientGAN and StackGAN++.

## 2. Additional IS Results

We also compute Inception score (IS) in different image resolution. At the resolution of  $128 \times 128$ , the IS of CookGAN is 6.1% better than ACME [1]. At the resolution of  $64 \times 64$ , CookGAN is better than  $R^2GAN$  [2] by 7.2%.

## References

- [1] Hao Wang, Doyen Sahoo, Chenghao Liu, Ee-peng Lim, and Steven CH Hoi. Learning cross-modal embeddings with adversarial networks for cooking recipes and food images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11572–11581, 2019.
- [2] Bin Zhu, Chong-Wah Ngo, Jingjing Chen, and Yanbin Hao. R2GAN: Cross-modal recipe retrieval with generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11477–11486, 2019.