S. No.	Paper Name	Main Takeaways
Score distillation Sampling Based Methods		
1	DreamFusion	1. You first generate image from NeRF then add noise, then make diffusion model predict the noise. The difference in noise is used to **update input while diffusion remains fixed** By doing the above you ensure that NeRF images match training distribution 2. Low-res images, slow to train (as NeRF is very tough to train), viewpoint not fine-grained (they use text like 'overhead view' to specify angles instead of absolute angles)
2	Magic3D	1. Fixes slow optimization and low-res problem 2. Uses coarse to fine structure for high res side (Two different diffusion models) 3. Uses **instant-NGP based representation** in place of vanilla NeRF to speed up coarse training, then uses, Latent diffusion in second stage for high-res image synthesis 4. Second stage uses **DMTet for updating the mesh directly**, and also has separate network for texture
3	3Dtopia	1. Extremely detailed captioning method. a. 2D captioning model for various renderings of images b. Combine captions with chatGPT c. Refine manually 2. Uses TriPlane representation for learning initial 3D meshes. THey also project triplane into 2D then into latent space to make the learning even faster

		3. In second stage it extracts mesh from triplane and converts it into SDF for better geometry learning, main logic is SDS is better for fine tuning and texture. 4. **This method is fast, has code, and uses 3D data, which we also have.**
4	ViewDiff	1. Can generate multiple consistent 3D views from text at once, these can be then used for Nerf training 2. Uses cross frame attention for consistency (its basically using attention across bunch of noise predictions in diffusion)and projection layer (kind of a complicated NerF like architecture) to ensure that object doesn't move around and remains in same position
5	Fantasia3D	1. Very similar to the fine part of Magic3D. DMTet for surface representation, and BRDF for material/texture. It also uses **normal map** as input to MLP unlike others that use images. 2. Since they use DMTet they can start with rough 3D shapes from users also 3. BRDF is just a MLP based material predictor
6	ProlificDreamer	1. Uses variational sampling on the NeRF side so that for same prompt you can sample from a distribution. 2. Basic idea is to have a whole bunch of theta's and run diffusion on all of them. The diffusion network can be LoRA or UNet 3. **Table 1 summarises dreamFusion, Magic3D, Fantasia3D and ProlificDreamer** in a succinct manner
7	DreamControl	1. I understood it the least 2. Tries to fix viewPoint bias and overfitting in first step. I understand the formula for viewPoint bias thing

		I have no idea how it is helping though. The overftting is avoided by using difference in boundary and surface density to define a good boundary and stop iterations after that. **3. Can generate animations.**
8	Progressive3D	1. Allows use of complex compositional prompts 2. Prompts are updated by selecting a region of interest where the model is updated. (This is done by using a 2D projection of a 3D bounding box) 3. Instead of retraining the model on a more complex prompt, they take projection of the noise on the source and target prompt to provide more useful guidance (**IMPORTANT IDEA**)
9	InterDreamer	1. Explores zero-shot motion generation. 2. While the project focuses on motion the idea for generating contacts from retrieval can be useful for our case.
10	InterFusion	1. Closest to our problem case. 2. Learns CLIP embedding to sample best poses for a prompt 3. Generates human and object with SDS in a combined fashion (**IMPORTANT IDEA**)
Main takeaways from all of above: 1) InstantNGP is a good representation to use if we plan to use NerF 2) Multistage architectures are working better than a single stage architecture 3) DMTet is better if we want to capture fine-grained details and its faster to train. (I don't understand why though)		
Efficient text-to-3D methods		
8	MobileGen3D	1. Uses NeRF to train a NeLF (instead of volume rendering it predicts color per ray). Since, NeLF has no volume rendering only rasterization it can speed up rendering. 2. To make customizable NeRF they train a NeRF then pass the renderings through diffusion, but

		they also share cross attention weights to maintain consistency. 3. Also does an extra step to sample pose uniformly **4. Only rendering is fast**
9	Text-to-3D Generative AI on Mobile Devices:	1. Good review of a few methods for how they perform on device. 2. Conclusion is that it is still an open problem.
Main Takeaways from efficient text-to-3D: 1) Currently there are no methods which focus on making the full pipeline efficient. 2) 3Dtopia still has the best training time. But it is only benchmarked on GPU not on device		
Direct 3D supervision method		
10	Point-e	1. Uses three step process, first text to image sampling, then image conditioned point cloud, the coarse point cloud conditioned fine point cloud **2. Relevant because we also have lots of 3D data** 3. We also have shap-e which works similarly.

Stable diffusion resources for future lab reference:

1) Best place to start: https://ayandas.me/2024/blog/diffusion-theory-from-scratch/
2) Lecture 12 and 13: https://www.youtube.com/playlist?list=PL2UML_KCiC0UPzjW9BjO-IW6dqliu9O4B
3) Classifier Free Guidance: https://sander.ai/2022/05/26/guidance.html