

Audio-Driven Image Generation

Snow



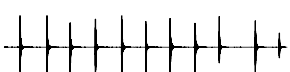
Wind noise



Splashing water



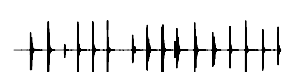
Ceramic



Gravel



Glass



Happy



Angry



Fearful

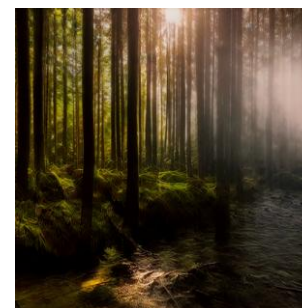
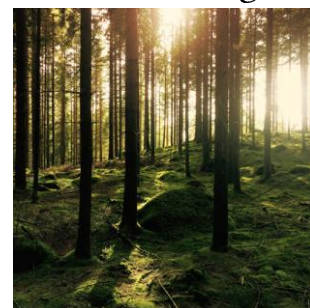


Audio-Driven Image Editing

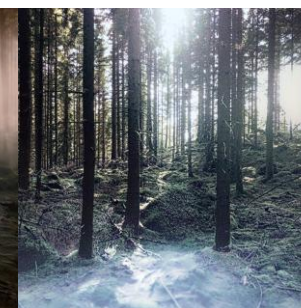
Squishing water



Source Image



Snow



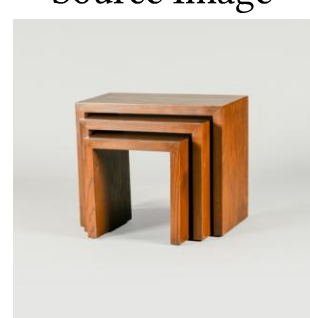
Rain



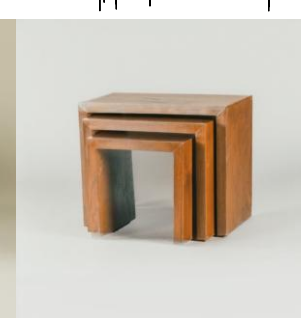
Metal



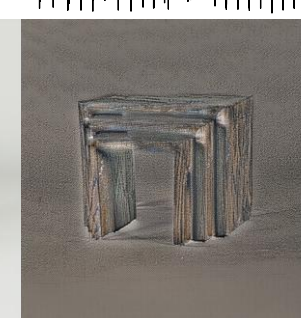
Source Image



Plastic Bag



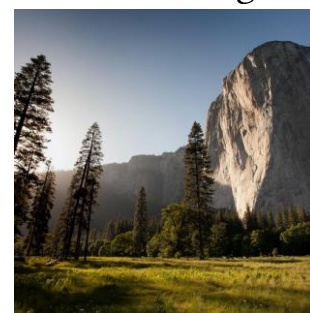
Gravel



Underwater bubbling



Source Image



Source Image



Happy



Miscellaneous

Generalizing to Out-of-Training Distribution Data

Source Image



Sad



Police Siren

Source Image



Synthesizing Stylized Images Using Both Text and Audio

Text Prompt:
"Pixel Art"

Audio Input:
Squishing water

