



Peak frame	Gender	Action Unit	Intensity
	0	6	3
		12	5
		25	5
		26	2

(a) Example of three categories for text-based dataset generation with peak image frame.

Action Unit	Description	AU subject	AU description 1	AU description 2
6	Cheek Raiser	cheek	raised	lifted
12	Lip Corner Puller	lip corner	pulled	stretched
25	Lips Part	lips	parted	separated
26	Jaw Drop	jaw	dropped	fallen

(c) Facial action units description for example facial image.

Text-based dataset modeling	A [Gender]’s [AU subject] is/are [AU intensity] [AU description], [Gender(pronoun)] ... 
Example	Ex) A man’s cheek is quite lifted, his lip corner is highly stretched, his lip is separated strongly, and his jaw is lightly dropped.

(b) Text-based dataset generation rule and example sentence.

1	2	3	4	5
marginally	slightly	kind of	positively	strongly
insignificantly	lightly	fairly	reasonably	highly
very slightly	somewhat	quite	sensibly	significantly

(d) Facial action unit intensities description. Each intensity is described as three different words with equivalent meaning.