Extended Data

Table A1 provides a comprehensive table of transformation guidelines for visual and descriptive changes based on weather conditions. It serves as a vital resource for creators and designers who aim to produce authentic and compelling weather-themed visuals or narratives. This chart categorizes instructions by weather type (e.g., rainy, snowy, foggy) and offers specific prompt conditions, sentence structures, keywords, and example outputs for each. By following these structured guidelines, users can achieve precise and meaningful adaptations for various weather scenarios. The table provides a clear guide for modifying images to represent different weather scenarios, ensuring consistent and contextually accurate changes in visual outputs or descriptive narratives.

Table A2 presents a detailed table outlining transformation guidelines for visual and descriptive modifications based on time and human attributes, such as gender, age, and skin tone. This table is an essential resource for ensuring accurate and inclusive representations across diverse scenarios. This chart categorizes instructions by specific conditions and provides sentence structures, keywords, and example outputs for each attribute or time-related change. Offering clear examples and structured prompts helps users navigate the complexities of representing temporal and personal attributes effectively. It serves as a reference for adjusting images to reflect various temporal settings (e.g., day, night) and individual characteristics, ensuring consistent and accurate transformations that align with specified attributes or times of day.

Table A3 provides a transformation matrix for sentence rewriting based on control words, illustrating methods to alter sentences by changing control words into various forms, such as verbs, adjectives, or nouns. This matrix serves as a practical guide for crafting dynamic and versatile sentence structures tailored to specific contexts or stylistic needs. Showcasing a variety of transformation examples helps users understand how subtle changes in control words can significantly impact sentence tone and meaning. This tool is especially valuable for content creators, linguists, and AI applications, offering a systematic approach to enhancing textual adaptability.

In our model, we refined input processing by splitting the text into individual words and creating targeted perturbations based on these words. This approach enhances granularity in understanding how each keyword influences the output. Fig. A1 illustrates how different models respond to prompts containing specific keywords. For example, with the keyword “snowing”, the generated image accurately depicts a snowy scene. However, when the keyword is omitted, the image reflects alternative visual outcomes. This highlights the model’s sensitivity to precise textual cues and the importance of keyword selection in achieving desired results.

This method enables us to assess the direct impact of individual keywords on image generation, revealing how each word shapes the output. By identifying and quantifying these effects, we can refine the model to enhance consistency and accuracy in generating context-specific visuals. By examining these effects, we can further optimize the model to ensure that subtle changes in textual input lead to precise and predictable alterations in the generated images. This methodology is a key step toward achieving robust and reliable AI-generated content.

Fig. A2 shows how an image changes when different prompt structures are used. This visualization underscores the pivotal role of prompt design in guiding AI-generated outputs. By tweaking keywords and phrases like “make it snowy” or “the weather must be snowy,” we see noticeable differences in how the image looks. Such variations illustrate the importance of nuanced prompt crafting to achieve specific visual goals. The significance of each word in the prompts is reflected by its weight, with darker colors representing words that have a more substantial influence. This weighted representation provides valuable insights into the linguistic components that drive visual transformations.

This analysis demonstrates the model’s sensitivity to specific prompt elements and highlights how prompt phrasing directly impacts the generated image’s characteristics. By leveraging this understanding, users can refine their prompts to produce more accurate and contextually relevant outputs, making this approach a powerful tool for optimizing AI-driven image generation.

Fig. A3 presents the heatmap analysis of prompt words to illustrate the influence of each word on the generated images using the Instruct-Pix2Pix model. This analysis provides a visual breakdown of how the model decodes and assigns importance to textual inputs. Each image corresponds to a specific prompt, with heatmaps showing the weight assigned to each word. Darker colors represent words with higher influence, indicating their more substantial impact on the resulting image transformation. This allows users to pinpoint which words are most critical for achieving desired visual outcomes.

This analysis highlights how the model interprets and prioritizes individual words in the prompt, allowing us to observe the relationship between textual emphasis and visual modifications. Such insights are invaluable for refining prompt engineering strategies, ensuring more precise and predictable image transformations. By understanding these dynamics, we can further optimize the Instruct-Pix2Pix model for enhanced sensitivity to nuanced textual cues, paving the way for improved AI-generated content.

Fig. A4, A5 illustrates the consistency of word influence in a single prompt applied across different images. This demonstrates the robustness of the models in maintaining predictable word-driven transformations across varying scenarios. The heatmaps display the weights assigned to each word in the prompt, with darker colors signifying higher influence. Such visual representation helps in understanding how individual words dominate the transformation process. The keyword “snowing” has a strong impact, resulting in reliable visual changes that align with the modification. This highlights the effectiveness of the keyword in directing the model’s focus toward the desired weather condition.

This analysis shows how the Instruct-Pix2Pix and Img2Img-Turbo models interpret the same prompt across various contexts.

Table A1: Weather Condition Transformation and Corresponding Keywords.

Condition	Specific Condition	Sentences	Keyword	Our Control
Rainy	Short Sentence	Transform the drizzle into a heavy downpour.	Transform, drizzle, heavy downpour	Heavy downpour
	Short Sentence	Transform the image to a rainy scene with dark clouds, steady raindrops, and puddles forming on the ground.	Rainy scene, dark clouds, steady raindrops, puddles	Rainy scene
	Short Sentence	Change the sunny weather to rainy, with dark clouds filling the sky and raindrops falling steadily.	Change, sunny weather, rainy, dark clouds, filling the sky, raindrops, falling steadily	Rainy, dark clouds, raindrops falling steadily
	Length Sentence	Turn the image to rainy and add a rainbow in the sky with a nice sense.	Turn, image, rainy, add, rainbow, sky, nice sense	Rainy, rainbow
Snowy	Short Sentence	Transform the weather to snowy.	Transform, weather, snowy	Snowy
	Short Sentence	Change the calm snow to heavy.	Change, calm snow, heavy	Heavy snow
	Length Sentence	Change the bright day to a snowy scene, with snowflakes gently falling and the ground covered in a white blanket.	Change, bright day, snowy scene, snowflakes gently falling, ground covered, white blanket	Snowy scene, snowflakes, ground covered in a white blanket
Foggy	Short Sentence	Change the foggy weather to a clear day.	Change, foggy weather, clear day	Clear day
	Length Sentence	Transform the night image to depict a foggy evening where visibility of the road is poor.	Transform, night image, foggy evening, visibility of the road, poor	Foggy evening, poor visibility
	Length Sentence	Switch the rain-drenched scene to one filled with fog, where the rain stops and a thick mist takes over, reducing visibility.	Switch, rain-drenched scene, filled with fog, rain stops, thick mist, reducing visibility	Filled with fog, thick mist, reducing visibility
Cloudy	Short Sentence	Change the cloudy sky to a clear blue sky.	Change, cloudy sky, clear blue sky	Clear blue sky
	Short Sentence	Transform the image to depict a cloudy, dark evening.	Transform, image, depict, cloudy, dark evening	Cloudy, dark evening
	Length Sentence	Heavy clouds fill the sky with thunder rumbling and bright lightning bolts striking down, creating an intense and stormy atmosphere.	Heavy clouds, fill the sky, thunder rumbling, lightning bolts, striking down, intense stormy atmosphere	Intense, stormy atmosphere

Table A2: Table of Transformation Guidelines for Visual and Descriptive Changes Based on Time, and Human Attributes.

Condition	Specific	Sentences	Keywords	Our Control
Person	Gender	Change the old woman to a young man in the image who is wearing sunglasses.	Change, old woman, young man, wearing sunglasses	Young man, wearing sunglasses
		Change the male in the image with a high-fashion woman.	Change, male, high-fashion woman	High-fashion woman
	Skin Color	Modify the person's complexion to a deep black hue.	Modify, person's complexion, deep black hue	Deep black hue
		Turn the white-skinned person to a black-skinned person.	Turn, white-skinned person, black-skinned person	Black-skinned person
	Age	Change the person's age to look considerably older.	Change, person's age, considerably older	Considerably older
		Change the elderly man into a young boy while retaining his distinctive facial characteristics.	Change, elderly man, young boy, retaining, distinctive facial characteristics	Young boy, distinctive facial characteristics
Tools	Replace	Remove the man's hat in this foggy image.	Remove, man's hat, foggy image	No hat
		Replace the traffic light with a simple light.	Replace, traffic light, simple light	Simple light
		Add a car crash to the street.	Add, car crash, street	Car crash
		Replace the black car with a white one.	Replace, black car, white car	White car
		Take off the man's glasses and put them on the lady's eyes.	Take off, man's glasses, put them on, lady's eyes	Glasses, the lady
	Change Color	Convert the red bus to a yellow one.	Convert, red bus, yellow one	Yellow
Time	Day	Convert a nighttime scene into a bright, sunny daytime setting with clear blue skies.	Convert, nighttime scene, bright, sunny daytime setting, clear blue skies	Bright, sunny daytime, clear blue skies
		Alter the nighttime environment into a lively day, with the darkness lifting and the sunlight bringing the world to life.	Alter, nighttime environment, lively day, darkness lifting, sunlight, bringing the world to life	Lively day, sunlight
	Night	Convert it to night with some noise of lights.	Nighttime, noise of lights	Night, noise of lights
		Turn this beautiful image to night and turn on the lights.	Nighttime, turn on lights	Night, lights
		Turn it to night with snowy weather.	Night, snowy weather	Night, snow

Table A3: Transformation Matrix for Sentence Rewriting Based on Control Words.

Control Words	Cloudy	Dark	Snowy	Rain, Rainbow	Heavy Downpour	Black-skinned Person
Control Words as Verb	Transform the image to be clouded and darkened.	Transform the image to be clouded and darkened.	Transform the weather to snow.	Turn the image to raining and add a rainbow in the sky with a nice sense.	Transform the drizzle to downpour heavily.	Turn the white-skinned person to blacken.
Control Words as Adjective	Transform the image to depict a cloudy, dark evening.	Transform the image to depict a darkening, overcast evening.	Transform the weather to snowy.	Turn the image to rainy and add a rainbow-colored sky with a nice sense.	Transform the drizzle to make it heavily downpouring.	Turn the person to make them black-skinned.
Control Words as Noun	Transform the image to depict the dark of a cloudy evening.	Transform the image to depict the dark of a cloudy evening.	Transform the weather to snowiness.	Turn the image to rain and add a rainbow in the sky with a nice sense.	Transform the drizzle into a downpouring rain.	Turn the white-skinned person into a person with black skin.
Use Word with 'ing'	The image is depicting a cloudy, dark evening.	The image is depicting a cloudy, dark evening.	Transform the weather to snowing.	Turn the image to rainy and add a rainbowing sky with a nice sense.	Transform the drizzle into a heavy downpouring.	Turn the white-skinned person into a person with black-skinning.
Sentences with Make	Make the image depict a cloudy, dark evening.	Make the image depict a cloudy, dark evening.	Make the weather to snowy.	Make the image to rainy and add a rainbow in the sky with a nice sense.	Make the drizzle into a heavy downpour.	Make the white-skinned person into a person with black skinning.
Sentences with Must	The image must depict a cloudy, dark evening.	The image must depict a cloudy, dark evening.	The weather must be snowy.	The image must be turned to rainy, and a rainbow must be added in the sky with a nice sense.	The drizzle must be transformed into a heavy downpour.	The white-skinned person must be turned into a black-skinned person.

Fig. A1: Results generated by Instruct-Pix2Pix showing how the presence of the keyword "snowing".

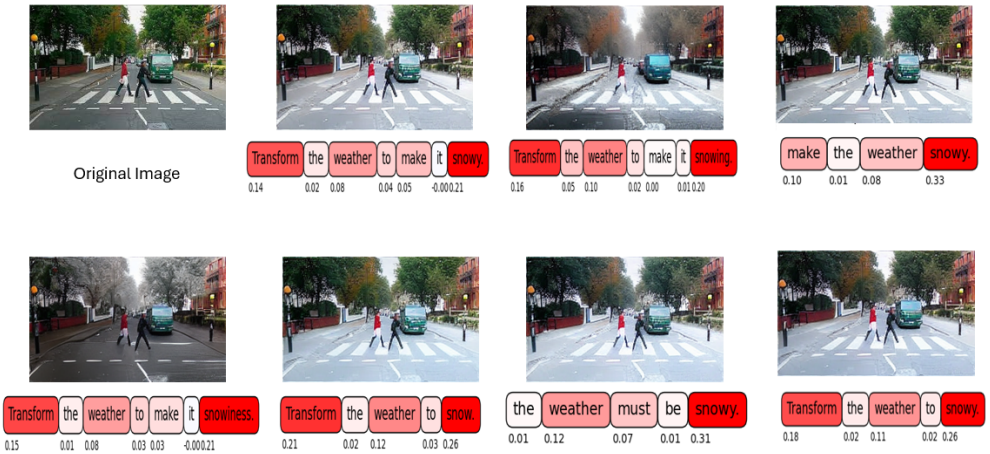

Fig. A2: Comparison of different structures of prompts used to generate the same image using Instruct-Pix2Pix.

Fig. A3: Heatmap analysis of prompts displaying the weights of individual words in guiding the Instruct-Pix2Pix model's visual output.

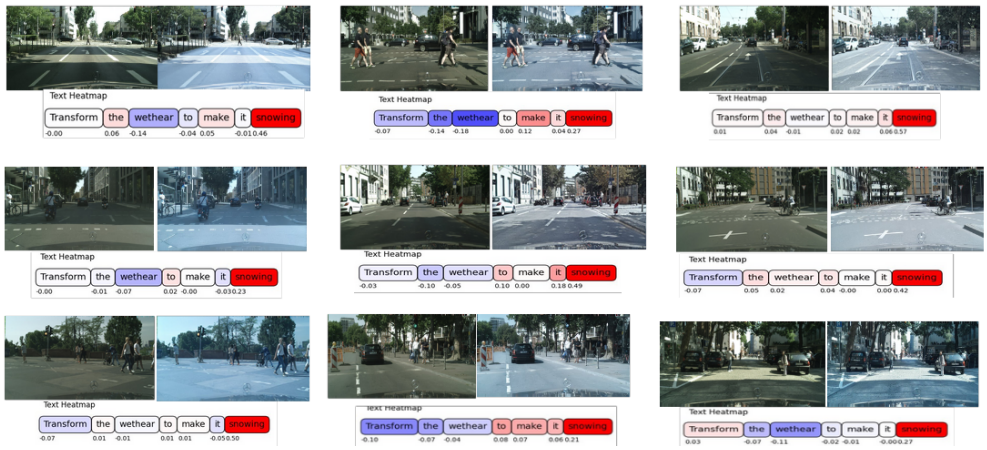

Fig. A4: Heatmap analysis of a single prompt applied across different images using Instruct-Pix2Pix.

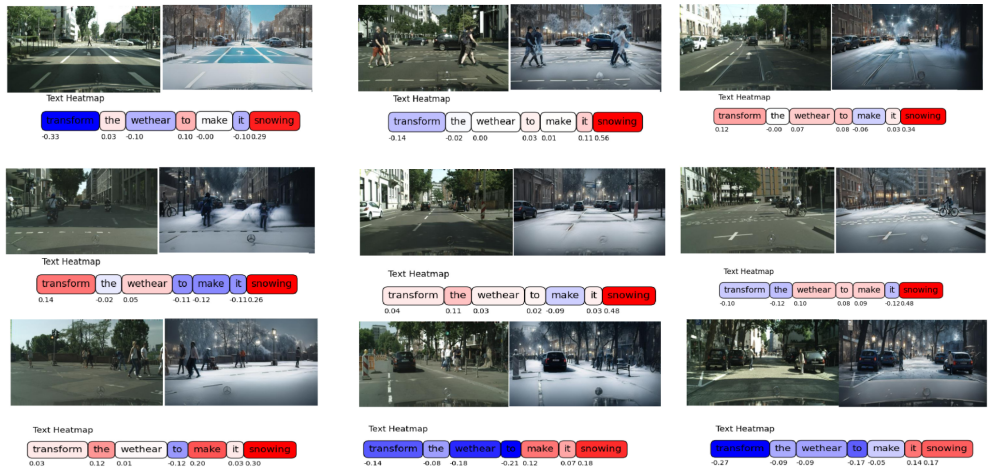

Fig. A5: Heatmap analysis of a single prompt applied across different images using Img2Img-Turbo.