Evaluation Experiment and Results

Intro to the experiment

Machine Learning (ML) is widely used in safety-critial systems such as autonomous vehicles, and to ensure safety, prior to deployment, they need to be tested against environment changes that could potentially lead to ML mistakes. Standard practice for testing is to simulate environmental changes using image transformations, such as reducing brightness with image transformation to simulate driving at night. However, the process of finding image transformations to simulate certain environment change is currently done intuitively and manually by experts. We want to propose a systematic and automatable approach that assists with this process. In order to evaluate the effectiveness of our approach, we are looking into conducting this experiment with experts in image processing to compare their manual results with our results and collect their feedback on our overall approach.

Procedure

We will collect both how participants identify image transformations simulating environment changes and their feedback on our proposed approach. Each participant will be compensated for their time. All participant information will be protected and each participant's responses will be identified only by number.

The participants first are introduced to the task, they are presented with necessary background on the 20 hazardous situations and 44 image transformations used during the experiment. Then the participants are asked to identify image transformations from the given list that can simulate each situation from the given list. We record the total time they take to complete this task. After this, we ask them to comment on the overall process and particularly what was challenging. They take a 20 mins break. Next, we present our systematic method to them and we discuss the difference in their and our results for identifying mistakes. Finally, we ask them to comment on how well we addressed the challenges they identified by selecting from the following choice list: 1) not addressed, 2) partially addressed, 3) fully addressed. The experiment with each participant is done individually. The results from participants are used to assess the accuracy of our proposed approach, and their overall feedback is concluded as evaluation of practicality of our approach.

Our experiment protocol has received ethics board approval.

Participant selection

Our inclusion criteria are individuals: 1) adults (18+) ; 2) who have obtained or are working towards a research graduate degree in computer science; 3) who have least one-year experience working directly in image processing; 4) who use English on a regular basis.
Our exclusion criteria are individuals who have direct conflict of interest to researchers on this team: 1) who are students or supervisors of anyone on the team; 2) who have business relations with anyone on the team; 3) members of this research team.

Individual mapping results

1.

<table>
<tr><td colspan="2" rowspan="2">Entry No. 3</td><td>Location, Guide Word, Parameter</td><td>Meaning</td><td>Consequence</td><td>Risk</td></tr>
<tr><td>Light Sources, Less (less of, lower), Number</td><td>Few light sources (fewer light sources than expected)</td><td>Too faint light (in parts of the scene)</td><td>Sensor will receive too faint light from some scene regions</td></tr>
<tr><td rowspan="5">Manual results</td><td>Participant 1</td><td colspan="4">RandomBrightnessContrast, RandomToneCurve</td></tr>
<tr><td>Participant 2</td><td colspan="4">PixelDistributionAdaptation + FDA + HistogramMatching (if the reference is dark), RandomBrightnesContrast, RandomGamma, RandomRain, ToGray, RandomShadow, ColorJitter, RandomToneCurve</td></tr>
<tr><td>Participant 3</td><td colspan="4">RandomBrightnessContrast</td></tr>
<tr><td>Participant 4</td><td colspan="4">ColorJitter, RandomBrightnessContrast, RandomRain, RandomSnow</td></tr>
<tr><td>Participant 5</td><td colspan="4">RandomBrightnessContrast</td></tr>
<tr><td colspan="2">Transformations agreed by the majority after discussion</td><td colspan="4">ColorJitter, RandomBrightnessContrast, RandomToneCurve, RandomRain</td></tr>
<tr><td rowspan="2">Systematic results</td><td>Effect + action</td><td>N/A</td><td>few light source; fewer light source</td><td>too faint light</td><td>sensor receive too faint some scene region light</td></tr>
<tr><td>Matched transformations</td><td>N/A</td><td>ColorJitter, RandomBrightnessContrast, RandomToneCurve</td><td>Same as Meaning</td><td>Same as Meaning</td></tr>
<tr><td>Automation results</td><td>Effect + action</td><td>N/A</td><td>few light source; fewer light source</td><td>too faint light parts of the scene</td><td>sensor receive too faint scene region light</td></tr>
</table>

	Matched transformations	N/A		ColorJitter, RandomBrightness Contrast	

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1 missed 2 transformations due to the challenge *many-to-many matching*.
Participant 2 included 6 more transformations due to *terminology mismatch*.
Participant 3 missed 3 transformations due to the challenge *many-to-many matching*.
Participant 4 missed 1 transformation due to the challenge *many-to-many matching*, included 1 more transformation due to wrong *filtering of relevant information*.
Participant 5 missed 3 transformations due to the challenge *many-to-many matching*.
As for our systematic approach, it only missed RandomRain, since this transformation has incomplete description. Our automation missed two transformations RandomRain (same reason) and RandomToneCurve (because similarity NLP models is missing background knowledge that experts have).

2.

Entry No. 86		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Light Sources, Reverse, Spectrum	The light source emits no light in the expected spectrum	Insufficient light in scene	Underexposure
Manual results	Participant 1	HueSaturationValue, MultiplicativeNoise, RandomToneCurve			
	Participant 2	PixelDistributionAdaptation + FDA + HistogramMatching (if the reference is dark),, RandomBrightnesContrast, RandomGamma, RandomRain, ToGray, ColorJitter, RandomToneCurve			
	Participant 3	RandomGamma			
	Participant 4	ColorJitter, RandomBrightnessContrast, RandomRain, RandomSnow			
	Participant 5	InvertImg, RandomBrightnessContrast			
Transformations agreed by the majority after discussion		ColorJitter, RandomBrightnessContrast, RandomToneCurve, RandomGamma			
Systematic results	Effect + action	N/A	light source emits no expected spectrum light	insufficient light	underexposure
	Matched transformations	N/A	ColorJitter, RandomBrightnessContrast,	Same as Meaning	

			RandomToneCurve		
Automation results	Effect + action	N/A	light source emits no spectrum light	Insufficient scene light	underexposure
	Matched transformations	N/A		ColorJitter, RandomBrightnessContrast	

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, missed 3 because of *many-to-many matching*, included 2 extra because of *terminology mismatch*.
Participant 2, included 5 extra because of *terminology mismatch*.
Participant 3, missed 3 because of *many-to-many matching*.
Participant 4, missed 3 because of *many-to-many matching*, included 1 extra because of *vague matching* and 1 other extra for *filtering relevant information*.
Participant 5, missed 3 transformations because of *many-to-many matching* and included 1 extra because of *terminology mismatch*.
As for our systematic approach, it missed RandomGamma because it relies on transformation description and the description for RandomGamma is missing in Albumentation. Our automation missed RandomToneCurve and RandomGamma for the same reasons: NLP model missing expert knowledge and missing description.

3.

Entry No. 124		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Light Sources, No (not none), Intensity	L.s. is off		Captured camera noise leads to fake effects
Manual results	Participant 1	ChannelDropout, InvertImg, Posterize, RGBShift, RandomBrightnessContrast, RandomToneCurve			
	Participant 2	PixelDistributionAdaptation and FDA (if the reference is dark), RandomBrightnesContrast, RandomGamma, ToGray, ChannelDropout, ColorJitter			
	Participant 3	ISONoise			
	Participant 4	GaussNoise, GlassBlur, ISONoise, MultiplicativeNoise			
	Participant 5	ISONoise, MultiplicativeNoise			
Transformations agreed by the majority after discussion		ColorJitter, RandomBrightnessContrast, ISONoise			

Systematic results	Effect + action	N/A	Light source is off		captured camera noise leads to fake effect
	Matched transformations	N/A	ColorJitter, Random-Brightness-Contrast		ISONoise
Automation results	Effect + action	N/A	Light source is off		captured camera noise leads to fake effect
	Matched transformations	N/A	ColorJitter, Random-Brightness-Contrast		ISONoise

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, missed 2 because of *many-to-many matching*, included 4 extra because of *terminology mismatch*.
Participant 2, missed 1 because of *many-to-many matching*, included 5 extra because of *terminology mismatch*.
Participant 3, missed 2 because of *many-to-many matching*.
Participant 4, missed 2 because of *many-to-many matching*, included 3 extra because of *terminology mismatch*.
Participant 5, missed 1 transformation becasue of *many-to-many matching* and included one more because of *terminology mismatch*.
As for our systematic approach and its automation, it is the exactly the same results.

4.

Entry No. 141		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Light Sources, Less (less of, lower), Beam properties	Focused beam	Only fractions of objects will be lit	Large parts of scene may be dark
Manual results	Participant 1	RandomSunFlare			
	Participant 2	PixelDistributionAdaptation + FDA + HistogramMatching (if the reference is dark), solarize, RandomShadow, ChannelDrop, RandomBrightnesContrast, RandomGamma, RandomRain, RandomToneCurve			
	Participant 3	RandomToneCurve			

	Participant 4	ColorJitter, RandomBrightnessContrast, RandomRain, RandomSnow			
	Participant 5	RandomSunFlare			
Transformations agreed by the majority after discussion		RandomBrightnessContrast, RandomToneCurve			
Systematic results	Effect + action	N/A	Focused beam	Only fractions of objects lit	dark large part of scene
	Matched transformations	N/A			RandomToneCurve
Automation results	Effect + action	N/A	Focused beam	Only fractions of objects lit	dark large part of scene
	Matched transformations	N/A			RandomToneCurve

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, missed 2 because of *terminology mismatch*.
Participant 2, included 8 extra because of *terminology mismatch*.
Participant 3, missed 1 because of *terminology mismatch.*
Participant 4, missed 1 because of *many-to-many matching*, included 2 extra because of *terminology mismatch* and 1 other extra for *filtering relevant information*.
Participant 5, missed 2 transformations because of *many-to-many matching*, included 1 extra because of *terminology mismatch*.
As for our systematic approach, as well as our automation, missed RandomBrightnessContrast because from the transformation description, it is not clear whether it can only change brightness of specific areas of the image.

5.

Entry No. 199		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Medium, Less (less of, lower), Spectrum	Medium acts as a filter and changes the light's spectrum	Skewed colours	Failed detection - since training data was obtained under unfiltered light
Manual results	Participant 1	ChannelShuffle, InvertImg, HueSaturationValue, Posterize, RBGShift, RandomToneCurve, ToSepia			
	Participant 2	ChannelDropout, ChannelShuffle, ColorJitter, PixelDistributionAdaptation,			

		FDA, HistogramMatching, FancyPCA, CLAE, HueSaturationValue, InvertImg, Posterize, RGBShift, RandomGamma, RamdomBrightnessContrast, RandomToneCurve, ToGray, ToSepia			
	Participant 3	RGBShift			
	Participant 4				
	Participant 5	ColorJitter, ChannelShuffle, RandomToneCurve			
Transformations agreed by the majority after discussion		RGBShift, ColorJitter, HueSaturationValue			
Systematic results	Effect + action	N/A	Medium acts as filter; Medium change light spectrum	skewed colours	
	Matched transformations	N/A	ColorJitter, Hue Saturation Value RGBShift	Same as meaning	
Automation results	Effect + action	N/A	Medium acts as filter; Medium change light spectrum	skewed colours	
	Matched transformations	N/A		ColorJitter	

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, missed 1 because of *many-to-many matching*, included 5 because of *terminology mismatch*.
Participant 2, included 13 extra because of *terminology mismatch*.
Participant 3, missed 2 because of *terminology mismatch*.
Participant 4, missed 3 because of *terminology mismatch*.
Participant 5, missed 2 because of *many-to-many matching*, and included 2 more because of *terminology mismatch*.
As for our systematic approach, found all the transformations. Our automation missed RGBShift because the model lacks expert knowledge to understand RGB is color.

6.

Entry No. 200	Location, Guide Word, Parameter	Meaning	Consequence	Risk
	Medium, As well as,	Medium has similar colour	Low contrast	Objects and medium

		Spectrum	as nearby l.s./object		become indis-tinguishable
Manual results	Participant 1	CLAHE, ColorJitter, Equalize, Normalize, RandomToneCurve			
	Participant 2	RandomGamma, RandomBrightnessContrast, ChannelDropout, Posterize, ToGray, ToSepia			
	Participant 3	RandomBrightnessContrast			
	Participant 4	ColorJitter, RandomBrightnessContrast			
	Participant 5	GaussianBlur, MedianBlur, Blur, AdvancedBlur, GlassBlur, Posterize			
Transformations agreed by the majority after discussion		ColorJitter, RandomBrightnessContrast, HueSaturationValue			
Systematic results	Effect + action	N/A	medium has similar colour as nearby light source; medium has similar colour as nearby light object	low contrast	Objects become indistinguish-able; medium become indis-tinguishable
	Matched transformations	N/A	ColorJitter, Hue Saturation Value	Random Brightness Contrast	
Automation results	Effect + action	N/A	medium has similar colour as nearby light source; nearby light object	low contrast	Objects become indistinguish-able; medium become indis-tinguishable
	Matched transformations	N/A		ColorJitter, Random-Brightness-Contrast, Downscale	

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, missed 2 because of *many-to-many matching*, included 4 more because of *terminology mismatch*.
Participant 2, missed 2 because of *many-to-many matching*, included 5 extra because of *terminology mismatch*.
Participant 3, missed 2 because of *terminology mismatch*.
Participant 4, missed 1 because of *many-to-many matching*,
Participant 5, missed 3 and included 5 more because of *filtering relevant information*.

As for our systematic approach, found all the transformations. Our automation missed HueSaturationValue because the model lacks expert knowledge to understand RGB is color.

7.

Entry No. 204		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Medium, More (more of, higher), Texture	Medium is more texturized than expected	Object appearance can be sig-nificantly distorted, e.g. fragmented	Object recognition hampered
Manual results	Participant 1	Emboss, GlassBlur, MotionBlur			
	Participant 2	AdvancedBlur, Blur, GaussNoise, GaussBlur, GlassBlur, ISONoise, MultiplicativeNoise, UnsharpMask, Median Blur, Morion Blur, Superpixels			
	Participant 3				
	Participant 4	GaussNoise, GlassBlur, ISONoise, MultiplicativeNoise			
	Participant 5	Sharpen, RingingOvershoot			
Transformations agreed by the majority after discussion		Description of this entry is too generic for matching transformations			
Systematic results	Effect + action	N/A	more texturized medium	significantly distorted object appearance	
	Matched transformations	N/A			
Automation results	Effect + action	N/A	more texturized medium	object appearance significantly distorted	
	Matched transformations	N/A			

8.

Entry No. 211		Location, Guide Word, Parameter	Meaning	Consequence	Risk

		Medium, Other than, Texture	Texture of medium is different than expected	Influences of medium on scene appearance is unexpected	Algorithms trained for a certain medium are confused
Manual results	Participant 1	Almost all transformations			
	Participant 2	Any transform except: Downscale, FromFloat			
	Participant 3				
	Participant 4				
	Participant 5	Sharpen, RingingOvershoot, PixelDistributionAdaptation, HistogramMatching, FDA			
Transformations agreed by the majority after discussion		Description of this entry is too generic for matching transformations			
Systematic results	Effect + action	N/A	different texture of medium	unexpected influences of medium on scene appearance	
	Matched transformations	N/A			
Automation results	Effect + action	N/A	different texture of medium	medium Influences appearance of scene	
	Matched transformations	N/A			

9.

Entry No. 371		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Object, More (more of, higher), Complexity	Object is more complex than expected		Parts of object not recognized or misinter-preted
Manual results	Participant 1	Emboss, GaussNoise, ISONoise, RandomRain, RandomSnow, Sharpen, Superpixels, Unsharpmask			
	Participant 2				
	Participant 3				

	Participant 4				
	Participant 5	RandomFog, RandomRain, RandomSnow, RandomShadow, RandomSunFlare, Superpixels			
Transformations agreed by the majority after discussion		Description of this entry is too generic for matching transformations			
Systematic results	Effect + action	N/A	more complex object		
	Matched transformations	N/A			
Automation results	Effect + action	N/A	Object more complex		
	Matched transformations	N/A			

10.

Entry No. 421		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Object, No (not none), Spectrum	Object has no specific colour	Object's colour fully depends on Incoming light, i.e. it reflects all light as received	Object is not recognized at all
Manual results	Participant 1	FDA, TemplateTransform			
	Participant 2	Solarize, RandomSnow			
	Participant 3	ToGray			
	Participant 4				
	Participant 5	ToSepia, ToGray, Solarize, RGBShift, InvertImg			
Transformations agreed by the majority after discussion		Description of this entry is too generic for matching transformations			
Systematic results	Effect + action	N/A	Object has no specific colour	Object colour fully depend on incoming light	

	Matched transformations	N/A	Color Jitter, Hue Saturation Value		
Automation results	Effect + action	N/A	Object has no specific colour	object colour fully depends on incoming light; i.e. reflects light	
	Matched transformations	N/A	Hue Saturation Value		

11.

Entry No. 478		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Object, More (more of, higher), Reflectance	Obj. has much Refl. (more than expected)	Shiny surface - mirror	Object not recognized
Manual results	Participant 1	FDA, TemplateTransform			
	Participant 2	Solarize, RandomSnow			
	Participant 3	ToGray			
	Participant 4				
	Participant 5	ToSepia, ToGray, Solarize			
Transformations agreed by the majority after discussion		No transformation can simulate this since no transformation is agreed by the majority of the participants			
Systematic results	Effect + action	N/A	Object has much relectance	shiny surface	
	Matched transformations	N/A			
Automation results	Effect + action	N/A	Object has much relectance	Shiny surface mirror	
	Matched transformations	N/A			

Particularly for this entry, compared to the results agreed by the majority of the participants:

Participant 1, included 2 more because of *terminology mismatch*.
Participant 2, included 2 extra because of *terminology mismatch*.
Participant 3, included 1 extra because of *terminology mismatch*.
Participant 5, included 3 extra because of *terminology mismatch*.
Our approach did not find any transformation.

12.

Entry No. 537		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Objects, No (not none), Number	Number of objects is not de-tectable/decidable	Scene with "unknown" number of objects	False negatives: CV alg. misses detection of some objects
Manual results	Participant 1	Almost all transformations			
	Participant 2	Blur, AdvancedBlur, GaussianBlur, MedianBlur, MotionBlur			
	Participant 3				
	Participant 4				
	Participant 5	RandomFog, RandomRain, RandomSnow, RandomShadow, RandomSunFlare			
Transformations agreed by the majority after discussion		Description is too generic			
Systematic results	Effect + action	N/A		unknown number of objects	
	Matched transformations	N/A			
Automation results	Effect + action	N/A		unknown number of objects	
	Matched transformations	N/A			

13.

Entry No. 611	Location, Guide Word, Parameter	Meaning	Consequence	Risk

		Objects, Less (less of, lower), Occlusion	Less objects occlude each other than expected	More details of objects are visible	Detection quality is decreased by too much clutter in the scene
Manual results	Participant 1	CLAHE, Emboss, RandomBrightnessContrast, Sharpen, UnsharpMask			
	Participant 2	Sharpen, UnsharpMask, Emboss			
	Participant 3				
	Participant 4	Emboss			
	Participant 5	RandomFog, RandomRain, RandomSnow, RandomShadow, RandomSunFlare			
Transformations agreed by the majority after discussion		Emboss, Sharpen			
Systematic results	Effect + action	N/A	less object occlude other	more of object detail visible	
	Matched transformations	N/A			
Automation results	Effect + action	N/A	less object occlude other	more object detail visible	
	Matched transformations	N/A			

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, included 3 more because of *terminology mismatch*.
Participant 2, included 1 extra because of *terminology mismatch*.
Participant 3, missed 2 because of *terminology mismatch*.
Participant 4, missed 1 because of *many-to-many matching*,
Participant 5, missed 2 and included 5 more because of *terminology mismatch*.
As for our systematic approach and automatic approach missed both transformations because the descriptions do not contain enough information,

14.

Entry No. 651	Location, Guide Word, Parameter	Meaning	Consequence	Risk
	Objects, More (more of, higher), Shadowing	More shadowing than expected	Large parts of scene in shadow	Underexposure: objects in shadow not detected

Manual results	Participant 1	Posterize, RandomShadow, RandomSunFlare, Solarize			
	Participant 2	RandomShadow, RandomBrightnessContrast, RandomGamma, PixelDistributionAdaptation, FDA, HistogramMatching			
	Participant 3	RandomGamma			
	Participant 4	RandomRain, RandomShadow			
	Participant 5	RandomShadow, RandomBrightnessContrast			
Transformations agreed by the majority after discussion		RandomShadow, RandomGamma			
Systematic results	Effect + action	N/A	More shadowing	Large shadow parts of scene	underexposure
	Matched transformations	N/A	RandomShadow	RandomShadow	
Automation results	Effect + action	N/A		Large shadow parts of scene	underexposure
	Matched transformations	N/A		RandomShadow	

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, missed 1 because of *many-to-many matching*, included 3 more because of *terminology mismatch*.
Participant 2, included 4 extra because of *terminology mismatch*.
Participant 3, missed 1 because of *many-to-many matching*.
Participant 4, missed 1 because of *many-to-many matching*, included 1 more because of *terminology mismatch*.
Participant 5, missed 1 because of *many-to-many matching*, included 1 more because of *terminology mismatch*.
As for our systematic approach and automation, both missed RandomGamma because its description is missing.

15.

Entry No. 827	Location, Guide Word, Parameter	Meaning	Consequence	Risk
	Observer - Optomechanics,	Orientation/Position is Other than	Tilt view	Objects not recognized

		Other than, Field of View	expected		
Manual results	Participant 1	MotionBlur, TemplateTransform			
	Participant 2				
	Participant 3				
	Participant 4	AdvancedBlur			
	Participant 5	AdvancedBlur, TemplateTransform			
Transformations agreed by the majority after discussion		No transformation in the given list can simulate this entry because no transformation is agreed by the majority of the participants.			
Systematic results	Effect + action	N/A	other Orientation/Position	tile view	
	Matched transformations	N/A			
Automation results	Effect + action	N/A	other position; other orientation	tile view	
	Matched transformations	N/A			

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, included 2 more because of *terminology mismatch*.
Participant 4, included 1 more because of *relevant information filtering*.
Participant 5 included 2 more because of *terminology mismatch*.
As for our systematic approach and automation, both missed RandomGamma because its description is missing.

16.

Entry No. 869	Location, Guide Word, Parameter	Meaning	Consequence	Risk
	Observer - Optomechanics, Temporal aperiodic,	VOrient changes in a temporally aperiodic	Motion blur	Occasional Misinterpretation of scene

		Viewing orientation	manner		
Manual results	Participant 1	MotionBlur			
	Participant 2	MotionBlur			
	Participant 3	MotionBlur			
	Participant 4	Advancedblur, Blur , GaussianBlur, GlassBlur, MedianBlur, MotionBlur, UnsharpMask			
	Participant 5	MotionBlur			
Transformations agreed by the majority after discussion		MotionBlur			
Systematic results	Effect + action	N/A	viewing orientation changes in temporally aperiodic manner	Motion blur	
	Matched transformations	N/A		Motion blur	
Automation results	Effect + action	N/A	viewing orientation changes in temporally aperiodic manner	Motion blur	
	Matched transformations	N/A		Advancedblur, Blur , GaussianBlur, GlassBlur, MedianBlur, MotionBlur	

Particularly for this entry, compared to the results agreed by the majority of the participants: Participant 4, included 6 more because of *relevant information filtering*.
As for our systematic approach found the right transformation. But automation has false negatives.

17.

Entry No. 910	Location, Guide Word, Parameter	Meaning	Consequence	Risk
	Observer - Op-	The sensor optics is more	Overall scene intensity as	Overexposure, at least of

		tomechanics, More (more of, higher), Transparency	transparent than expected	received by the electronics is higher than expected	specific objects
Manual results	Participant 1	ColorJitter, HueSaturationValue, RGBShift, RandomBrightnessContrast, RandomSunFlare, RandomToneCurve			
	Participant 2	RandomGamma, RandomBrightnessContrast, ColorJitter			
	Participant 3	RandomGamma			
	Participant 4	Color Jitter, RandomBrightnessContrast, RandomRain, RandomSnow			
	Participant 5	Normalize, RandomSunFlare, HueSaturationValue, Equalize, RandomBrightnessContrast, RandomOvershoot			
Transformations agreed by the majority after discussion		ColorJitter, RandomGamma, RandomBrightnessContrast			
Systematic results	Effect + action	N/A	more transparent sensor optics	higher Overall scene intensity	overexposure of specific object
	Matched transformations	N/A		ColorJitter, RandomBrightness Contrast	
Automation results	Effect + action	N/A	more transparent sensor optic	higher overall scene intensity; electronics	overexposure of specific objects
	Matched transformations	N/A		ColorJitter, RandomToneCurve	

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, missed 1 because of *many-to-many matching*, included 4 more because of *terminology mismatch*.
Participant 3, missed 3 because of *many-to-many matching*
Participant 4, missed 1 because of *many-to-many matching*, included 1 more because of *terminology mismatch* and included 1 more because of *relevant information filtering.*
Participant 5, missed 2 because of *many-to-many matching*, included 5 more because of *terminology mismatch*
As for our systematic approach missed RandomGamma because of the missing description.
Automation missed 2 and included one more because of the limitation of knowledge.

18.

Entry No. 1017		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Observer - Optomechan-ics, Less (less of, lower), Focusing	DoF is smaller than expected	Blurry image	Detection of edges and their correct position deteriorated
Manual results	Participant 1	AdvancedBlur, Blur, Equalize, GaussianBlur, GlassBlur, MedianBlur, MotionBlur, Normalize, RandomFog			
	Participant 2	Blur, AdvancedBlur, GaussianBlur, MedianBlur, MotionBlur			
	Participant 3	AdvancedBlur			
	Participant 4	AdvancedBlur, Blur, GaussianBlur, GlassBlur, MedianBlur, MotionBlur, RandomRain, UnsharpMask			
	Participant 5	GaussianBlur, MedianBlur, Blur, AdvancedBlur, GlassBlur			
Transformations agreed by the majority after discussion		Blur, AdvancedBlur, GaussianBlur, MedianBlur, MotionBlur, GlassBlur, Downscale			
Systematic results	Effect + action	N/A	smaller DoF	blurry image	
	Matched transformations	N/A		Blur, AdvancedBlur, GaussianBlur, MedianBlur, MotionBlur, GlassBlur, Downscale	
Automation results	Effect + action	N/A	smaller depth of field	blurry image	
	Matched transformations	N/A		AdvancedBlur, Blur, Downscale, GaussianBlur, GlassBlur, MedianBlur, MotionBlur, Sharpen, UnsharpMask	

Particularly for this entry, compared to the results agreed by the majority of the participants:
Participant 1, included 4 more because of *terminology mismatch*.
Participant 2, missed 1 because of *many-to-many matching*
Participant 3, missed 5 because of *many-to-many matching*
Participant 4, included 2 more because of *terminology mismatch*

Participant 5, missed 2 because of *many-to-many matching*.
As for our systematic approach, it found all the transformations. Our automation missed Downscale and included Sharpen and UnsharpMask because of the lack of CV/IP knowledge.

19.

Entry No. 1120		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Observer - Electronics, More (more of, higher), Exposure and shutte	Longer expo-sure time than expected	More light captured per image than expected	Overexposure
Manual results	Participant 1	ColorJitter, HueSaturationValue, RGBShift, RandomBrightnessContrast, RandomSunFlare, RandomToneCurve, Solarize			
	Participant 2	RandomGamma, RandomBrightnessContrast, ColorJitter			
	Participant 3	RandomGamma			
	Participant 4	ChannelDropout, ColorJitter, RandomBrightnessContrast			
	Participant 5	RandomBrightnessContrast, RandomSunFlare, RandomOvershoot, Normalize, HueSaturationValue			
Transformations agreed by the majority after discussion		RandomGamma, RandomBrightnessContrast, ColorJitter			
Systematic results	Effect + action	N/A	Longer expo-sure time	More light captured	Overexposure
	Matched transformations	N/A		Color Jitter, Random Brightness Contrast	
Automation results	Effect + action	N/A	Longer expo-sure time	More light	Overexposure
	Matched transformations	N/A		Color Jitter, Random Brightness Contrast	

Particularly for this entry, compared to the results agreed by the majority of the participants:

Participant 1, missed 1 because of *many-to-many matching*, included 5 more because of *terminology mismatch*.
Participant 3, missed 2 because of *many-to-many matching*
Participant 4, missed 1 because of *many-to-many matching*, included 1 more because of *terminology mismatch*.
Participant 5, missed 2 because of *many-to-many matching*, included 4 more because of *terminology mismatch*.
As for our systematic approach and automation, they found all the transformations.

20.

Entry No. 1159		Location, Guide Word, Parameter	Meaning	Consequence	Risk
		Observer - Electronics, Less (less of, lower), Resolution (spatial)	Pixel size is larger than expected	Pixel size is larger than expected	Lower noise level than expected
Manual results	Participant 1	Superpixels			
	Participant 2	Sharpen, UnsharpMask, Emboss			
	Participant 3	Downscale			
	Participant 4	Superpixels			
	Participant 5	Downscale, Superpixels, GaussianBlur, MedianBlue, Blur, GlassBlur			
Transformations agreed by the majority after discussion		Superpixels, Downscale			
Systematic results	Effect + action	N/A	larger pixel size	larger pixel size	lower noise level
	Matched transformations	N/A	Superpixels, Downscale	Same as Meaning	
Automation results	Effect + action	N/A	larger pixel size	larger pixel size	lower noise level
	Matched transformations	N/A	Downscale, To Gray	Same as Meaning	Glass Blur, ISO Noise

Particularly for this entry, compared to the results agreed by the majority of the participants:

Participant 1, missed 1 because of *many-to-many matching*.
Participant 2, missed 2 and included 3 because of *terminology mismatch*.
Participant 3, missed 1 because of *many-to-many matching*
Participant 4, missed 1 because of *many-to-many matching*.
Participant 5, included 4 because of *terminology mismatch*.
As for our systematic approach found all the transformations. Our automation missed superpixels and included 3 more because of its lack of CV/IP knowledge.

Participant comments on the mapping process

1. Participant 1
   Process: First read through all transformations and took notes on what each transformation is changing. Then they went over the CV-HAZOP entries list, highlight text they thought correspond to a transformation. Then map their notes of transformations to the highlighted text.
   Comments: The process has square complexity, it is really hard to not miss anything in the process. Some descriptions of transformations are not clear, it is hard to judge exactly what each transformation does. Some entries of CV-HAZOP is quite vague and it can be matched to pretty much all the transformations, like changing texture.

2. Participant 2
   Process: First take an entry, then check all the transformations descriptions to see if they can simulate this entry, then start understanding and using his knowledge to skip some transformations when the entry needs a specific scene change or whether a transformation is totally unrelated.
   Comments: Overall the process is long and annoying. The whole task is underspecified, it is not clear what exactly should be done and what criteria should be used to find the transformations since the transformations are not directly corresponding to the entries. The documentation for transformation are not good, hard to understand, input of function not well explained. Therefore, it is easy to misunderstand or forget some. The hardest part is to translate CV-hazop changes to how it should look in an image. Sometimes too many transformations can correspond to one CV-HAZOP entry.

3. Participant 3
   Process: First read through and understand the transformations, then go over each CV-HAZOP entry and find the best match of image transformation.
   Comments: the mapping is not one-to-one, also some transformations can be “hacked” with extreme values to simulate some entries. Some descriptions are too vague, it can be simulated in many different ways, also with composition of transformations. It would be useful if they had examples. The descriptions are also not written in proper CV terminologies which makes the mapping hard. The input descriptions need to be pre-processed to make the task easier. Also that the whole task is very vague, which makes it very hard to decide what is a good match.

4. Participant 4:
   <u>Process:</u> First read through the transformations, classified them into groups: noise, brightness, blurring, compression etc., then go over each CV-HAZOP entry and try to match groups of transformations into the entries.
   <u>Comments:</u> The process is very long and boring. You can easily miss some transformations for one entry. Some descriptions of transformations are misleading. In certain cases, it's hard to borderline transformations.

5. Participant 5:
   <u>Process:</u> First read through transformations and CV-HAZOP entries, try to understand what change is being described, then match the changes.
   <u>Comments:</u> The process is long and prone to mistakes. The mapping is not clear. The "in some scene regions" is confusing, since the transformations are mostly applying changes to the entire image. It is also confusing that sometimes some transformations can do more than what's in the entries. It is not clear whether this can be treated as separate transformations.

Participant feedback on our approach

1. Participant 1
   For many-to-many matching: it is fully addressed.
   For filtering relevant information: it seems to be fully addressed because they can't seem to come up with a counterexample.
   For terminology mismatch: fully addressed since automation does not need human judgement.

2. Participant 2
   They think it is useful to automate since it hugely reduces human effort.
   Is our approach a clearer process for performing the task?: If it works perfectly yes, but it is hard to make guarantees due to the complexity of the steps
   For many-to-many matching: Partially addressed because it is possible that two transformation could simulate one change by combining the effect that it could simulate.
   For terminology mismatch: Partially addressed. This is a good attempt to move forward towards addressing it, it's reasonable taking into account it's different language and the angle from which the changes takes place (are described), etc..

3. Participant 3
   It is impossible to fully make the task automated and accurate since everything is written in natural language and not formalized, this always creates mistakes.
   For many-to-many mapping: partially addressed
   For terminology mismatch: partially addressed since it depends on the performance of the ML model.

4. Participant 4
   For many-to-many matching: it can fully addressed if the input descriptions are ideal.
   For underspecified process, it can be fully addressed if one/NLP models have enough domain knowledge.
   For boring and long process, the automated approach is very useful.

5. Participant 5
   They think all challenges are fully addressed.
   They think the approach makes sense to then and is logical. The approach is also close to how they (or humans) might arrive at matching. But, the bottleneck here may be the albumentation API description quality and the CV-HAZOP list description. If the API description of the albumentation is enhanced by considering text from a google search of that transformation into the NLP model, they would imagine your approach to improve the matching quite close to what humans could have come up with (or atleast what they came up with). If boths inputs are more accurate, and more descriptive, the propose approach could scale and perform accuretly.