A Similarity Measure for Illustration Style
Supplemental Material #2 of 2: All Mash-Ups and Other Results

Elena Garces
Universidad de Zaragoza

Aseem Agarwala
Adobe Systems, Inc.

Diego Gutierrez
Universidad de Zaragoza

Aaron Hertzmann
Adobe Systems, Inc.

Additional Details on Learning and Evaluation

On pages 2-5, we show:

- The learned weight vector $\mathbf{w}$ for the whole set of features (Figure 1)

- The influence of varying the parameter λ in the error measures *accuracy* and *perplexity*

- 24 triplets where our metric disagree with Turkers (Figures 3 and 4)

Additional details and results from the Mash-up Evaluation

(Pages 6-8) We performed two tests: the first one was an open-ended task, where users were free to select background, the story and the number of images to use (see Figure 5). The second one was guided: we required at least 4 pieces of clip art in each composition, and we prevented the same user from working on the same story twice (see Figures 6 and 7).

Our pool of stories were the following:

- Show summer season at the lake. Consider including animals, plants, weather elements and/or human activities

- Show the world after an alien invasion

- You are looking at the sky through your window. Include elements you might see in the sky

- Create a poster for a gangster movie, or a scene of the movie

- Create the poster for a scary movie, or a scene of the movie

- Wildlife: Picture the jungle or a forest, and some interaction between animals

- Choose any character (keyword: "face") and show him/her experiencing one or more desires or feelings

- Illustrate the following situation: it is your birthday, and you are feeling very happy with all the presents

In both studies: time was limited to eight minutes; we paid $0.30 per task; and we offered a bonus of $1 for the best compositions.
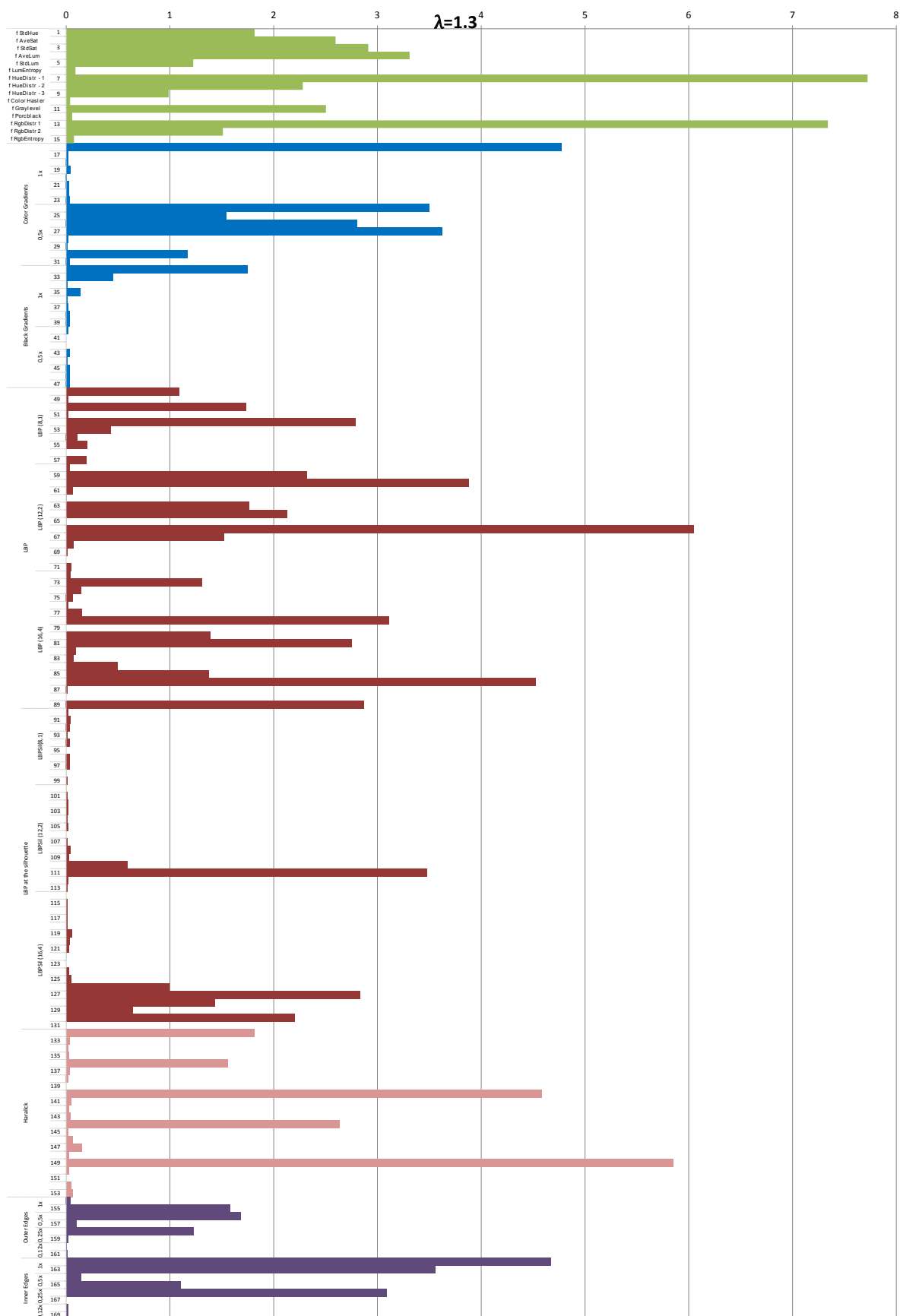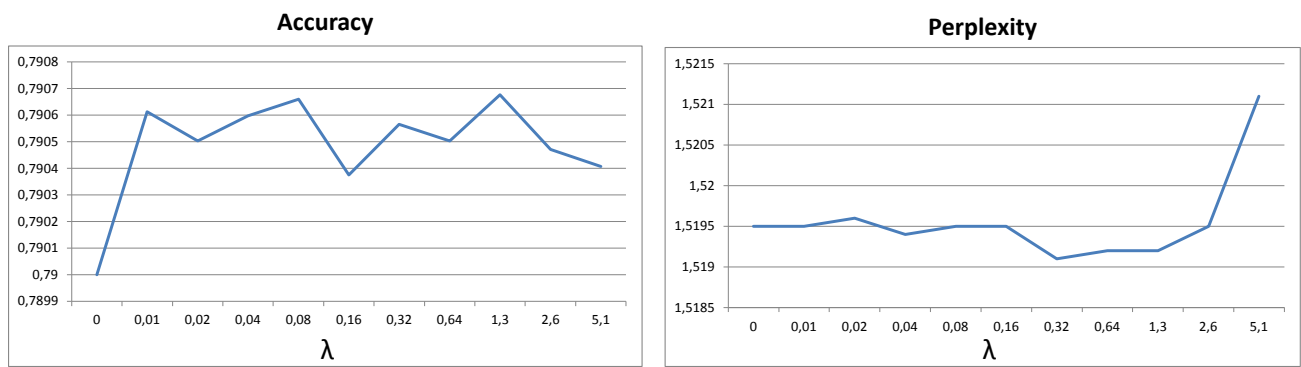
Figure 1: *Learned weights* **w**

Figure 2: *L1 Regularization: accuracy and perplexity for different values of lambda*

A	B	C	#AB	#AC	Agreement	Probability
			3	8	0.189	0.895
			10	1	0.19	0.164
			5	6	0.196	0.947
			9	1	0.201	0.172
			6	5	0.204	0.0569
			1	9	0.21	0.82
			5	6	0.214	0.937
			2	9	0.219	0.837
			2	9	0.221	0.835
			6	5	0.236	0.0756
			3	7	0.237	0.863
			4	6	0.246	0.896

Figure 3: *Testing triplets with greatest disagreement between Turkers and our learned similarity. The triplets are sorted by "agreement," which is equal to the probability of the Turk scores according to the learned model; agreement of 1 would be perfect agreement. #AB is the number of Turkers who rated image A as more similar to B than to C; #AC is the number of Turkers who rated A as more similar to C. The "probability" column shows the model's probability that A and B are more similar than B and C.*

A	B	C	#AB	#AC	Agreement	Probability
			1	9	0.252	0.777
			3	7	0.259	0.843
			3	7	0.262	0.841
			6	5	0.262	0.944
			12	0	0.269	0.269
			4	7	0.281	0.851
			9	2	0.281	0.224
			7	4	0.282	0.15
			8	2	0.284	0.22
			2	9	0.291	0.766
			5	6	0.292	0.884
			3	8	0.302	0.79

Figure 4: *The next 12 testing triplets with greatest disagreement between Turkers and our learned similarity. The triplets are sorted by "agreement," which is equal to the probability of the Turk scores according to the learned model; agreement of 1 would be perfect agreement. #AB is the number of Turkers who rated image A as more similar to B than to C; #AC is the number of Turkers who rated A as more similar to C. The "probability" column shows the model's probability that A and B are more similar than B and C.*
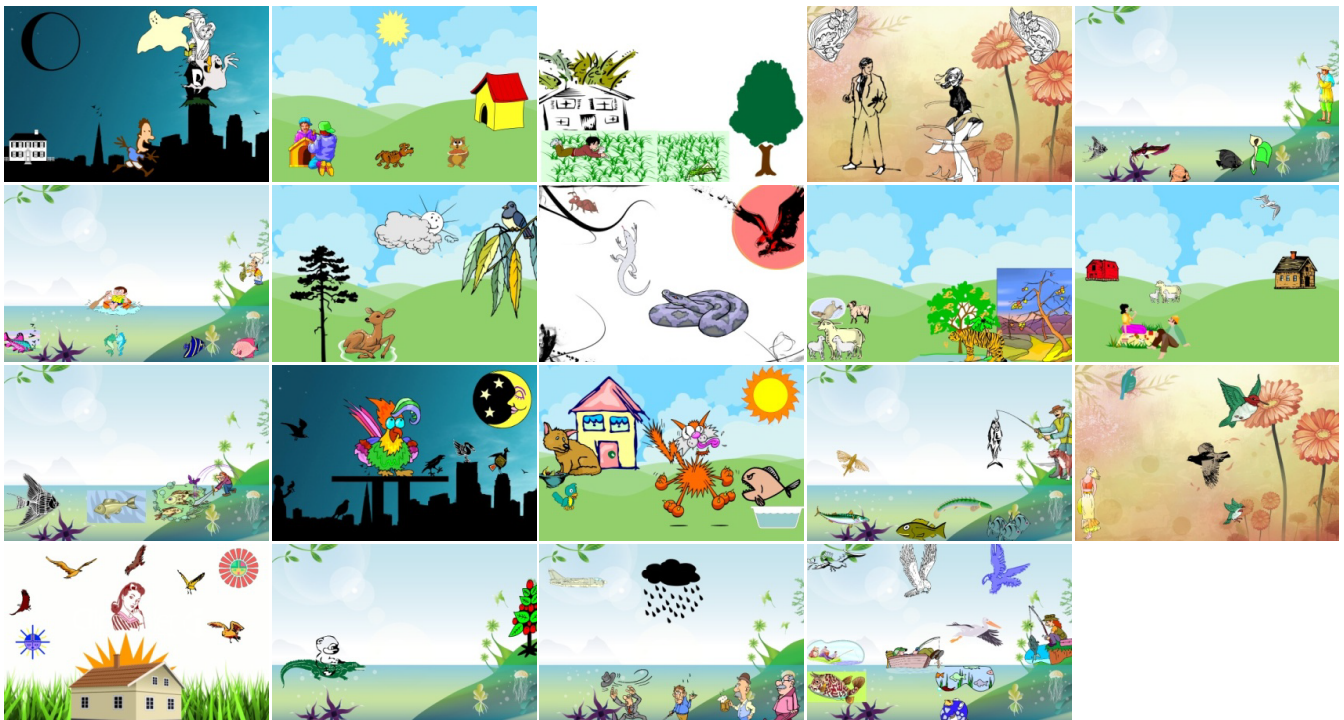
Figure 5: *Open-ended task: The upper block shows compositions created using the similarity metric. The lower block shows compositions created without it.*
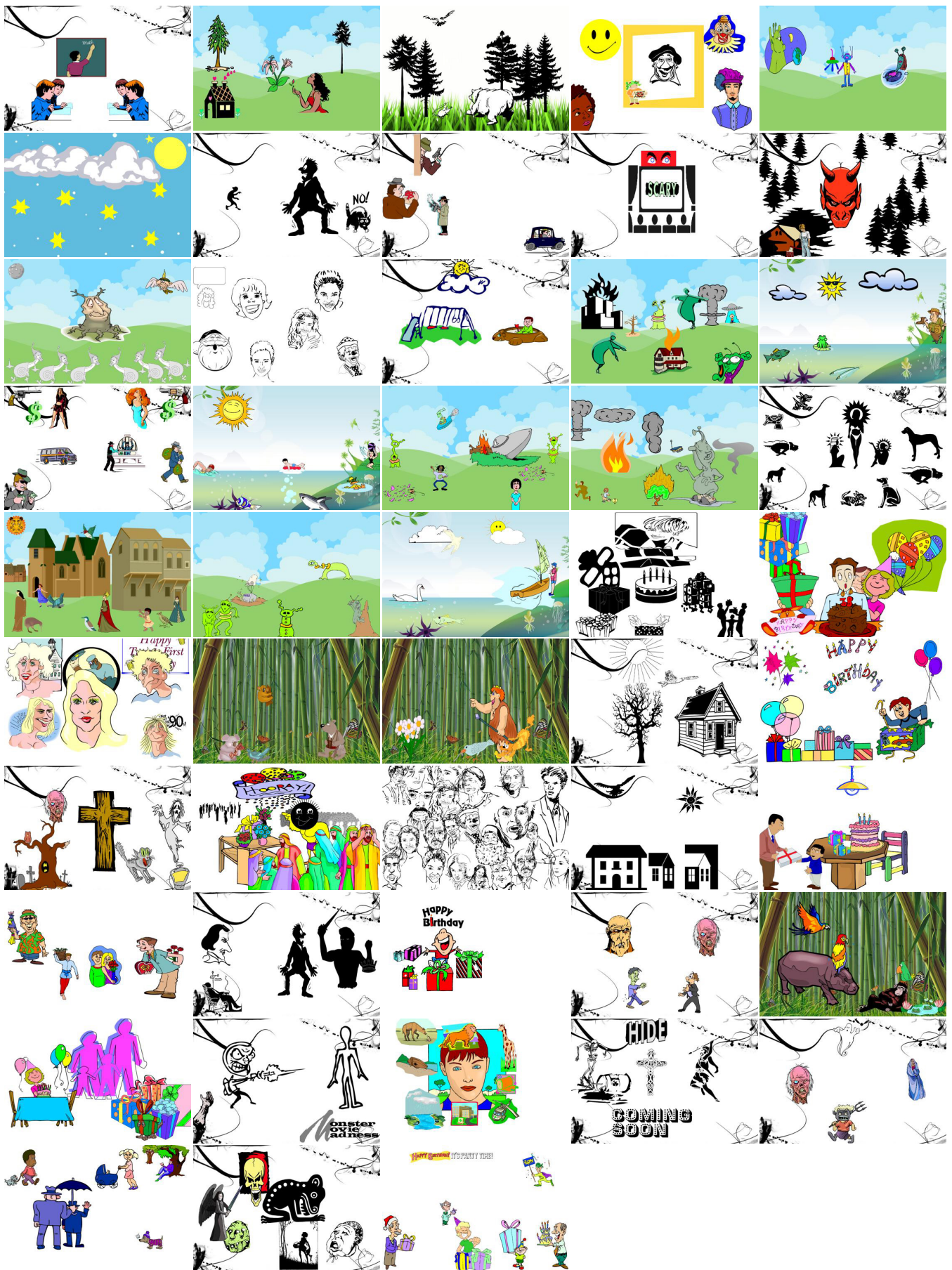
Figure 6: *Guided task: Compositions created with the similarity metric turned on.*
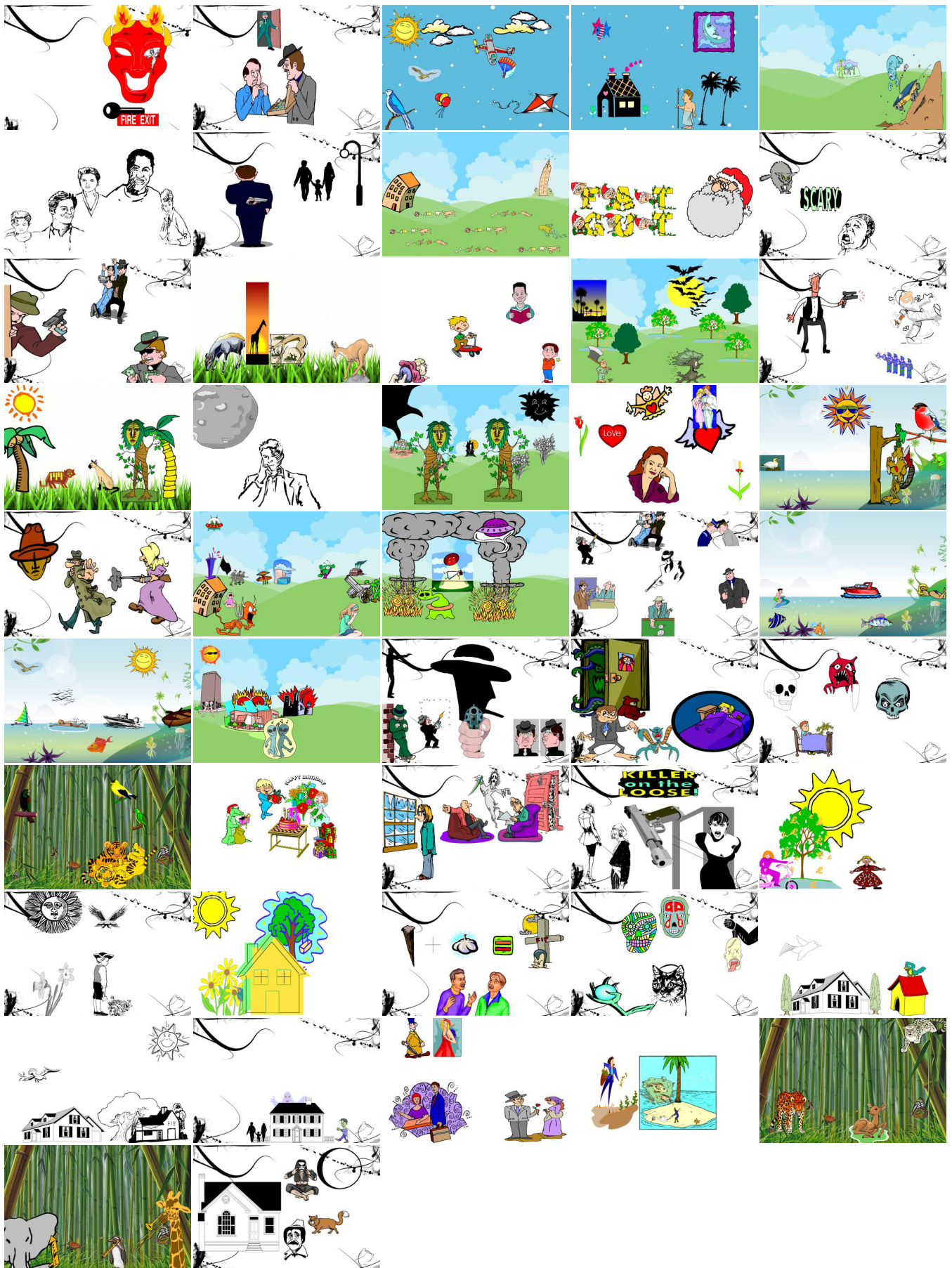
Figure 7: *Guided task: Compositions created with the similarity metric turned off.*