Perceptual metric for color transfer methods

H. Hristova O. Le Meur R. Cozot K. Bouatouch

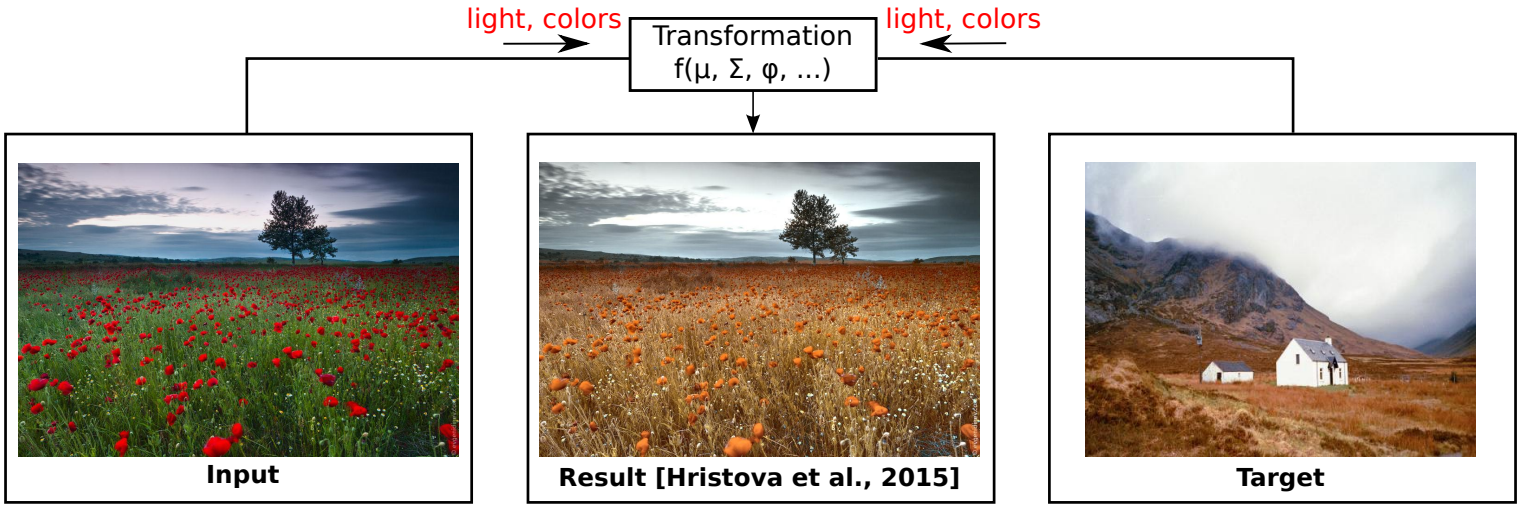
Outline

Outline

Color transfer

Overview

- Modification of the colors of an *input* image according to a *target* color palette;
- Statistical transformations between the input and target color distributions;
- Applications: image/video color grading [FPC*14, BSPP13], time-lapse image hallucination [SPDF13], etc.



Color transfer

Which one is best output?

- A number of plausible outputs;
- Lack of a ground truth.

Outline

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;
- Various protocols leading to biasness.

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;
- Various protocols leading to biasness.

Objective evaluation

A good color transfer should [FSDH14, HLMCB15]:

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;
- Various protocols leading to biasness.

Objective evaluation

A good color transfer should [FSDH14, HLMCB15]:

- preserve the integrity of the result, *i.e.* not introduce artifacts;

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;
- Various protocols leading to biasness.

Objective evaluation

A good color transfer should [FSDH14, HLMCB15]:

- preserve the integrity of the result, *i.e.* not introduce artifacts;
- transfer correctly the target color palette.

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;
- Various protocols leading to biasness.

Objective evaluation

A good color transfer should [FSDH14, HLMCB15]:

- preserve the integrity of the result, *i.e.* not introduce artifacts;
- transfer correctly the target color palette.

Objective metrics:

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;
- Various protocols leading to biasness.

Objective evaluation

A good color transfer should [FSDH14, HLMCB15]:

- preserve the integrity of the result, *i.e.* not introduce artifacts;
- transfer correctly the target color palette.

Objective metrics:

- SSIM(input,result);

Quality assessment of color transfer

Subjective user study

- Efficient way to study the subjectiveness;
- Time-consuming process;
- Various protocols leading to biasness.

Objective evaluation

A good color transfer should [FSDH14, HLMCB15]:

- preserve the integrity of the result, *i.e.* not introduce artifacts;
- transfer correctly the target color palette.

Objective metrics:

- SSIM(input,result);
- Bhattacharya coefficient(result,target).

Quality assessment of color transfer

Motivation

1. Independent objective metrics may be weak predictors of the color transfer quality.
2. For a stronger prediction model - an ensemble of features which account for subjective aspects of the color transfer evaluation.

Quality assessment of color transfer

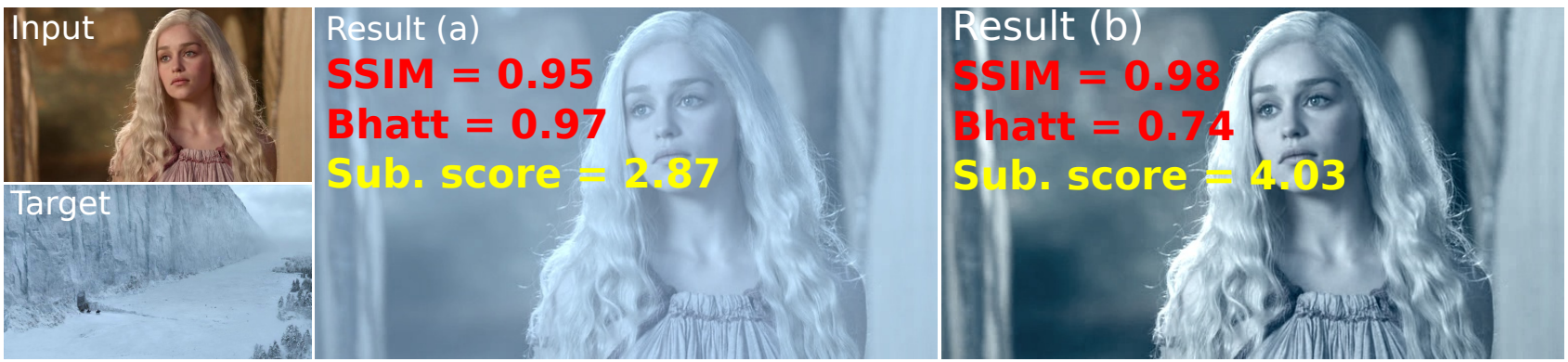
Motivation

1. **Independent objective metrics may be weak predictors of the color transfer quality.**
2. For a stronger prediction model - an ensemble of features which account for subjective aspects of the color transfer evaluation.

Quality assessment of color transfer

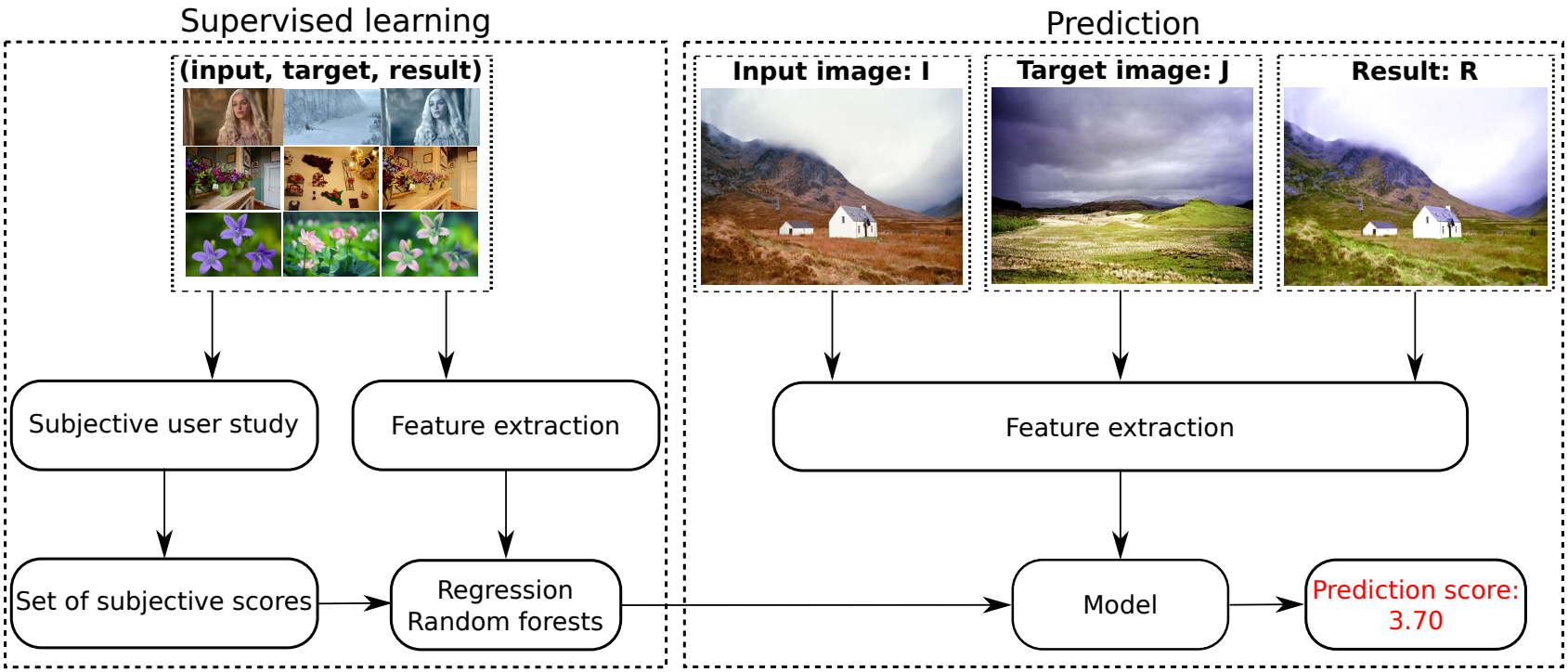
Motivation

1. Independent objective metrics may be weak predictors of the color transfer quality.
2. **For a stronger prediction model - an ensemble of features which account for subjective aspects of the color transfer evaluation.**

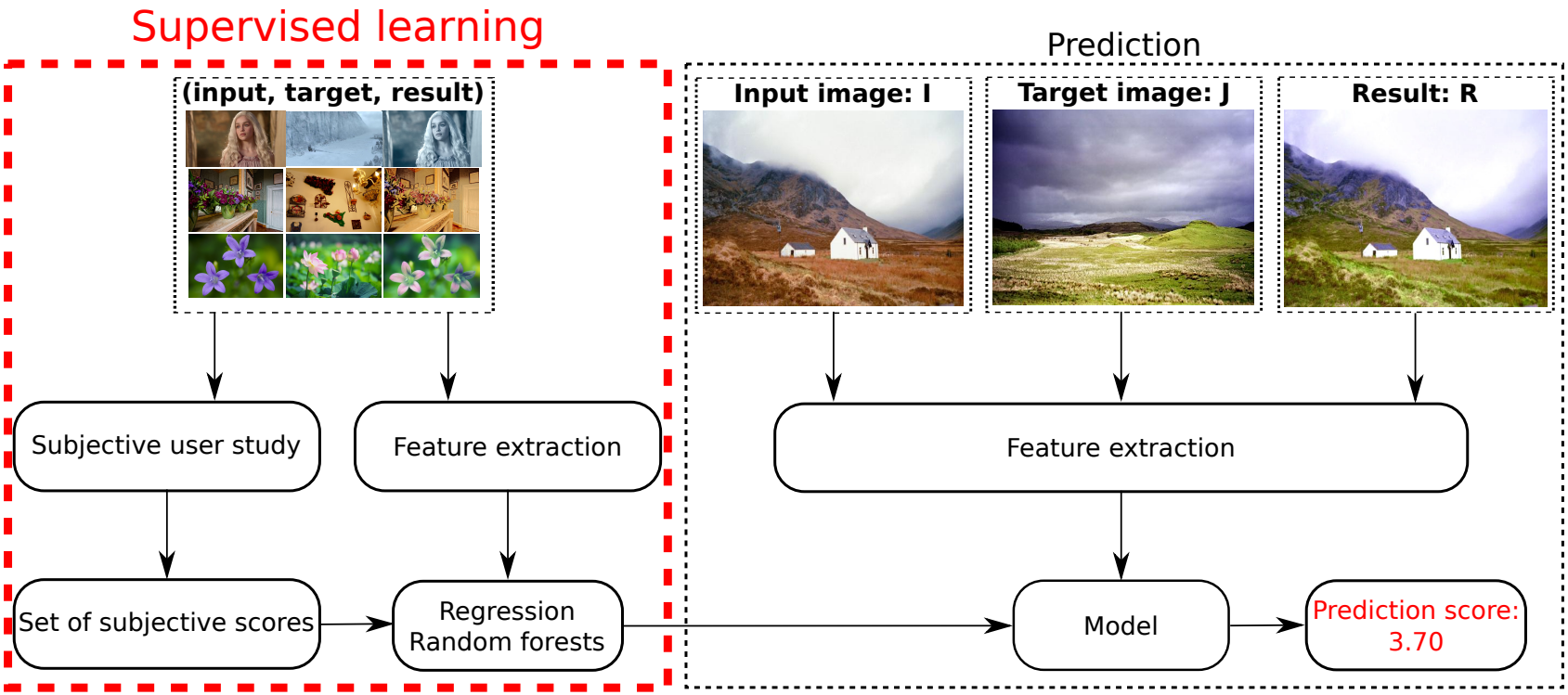


Outline

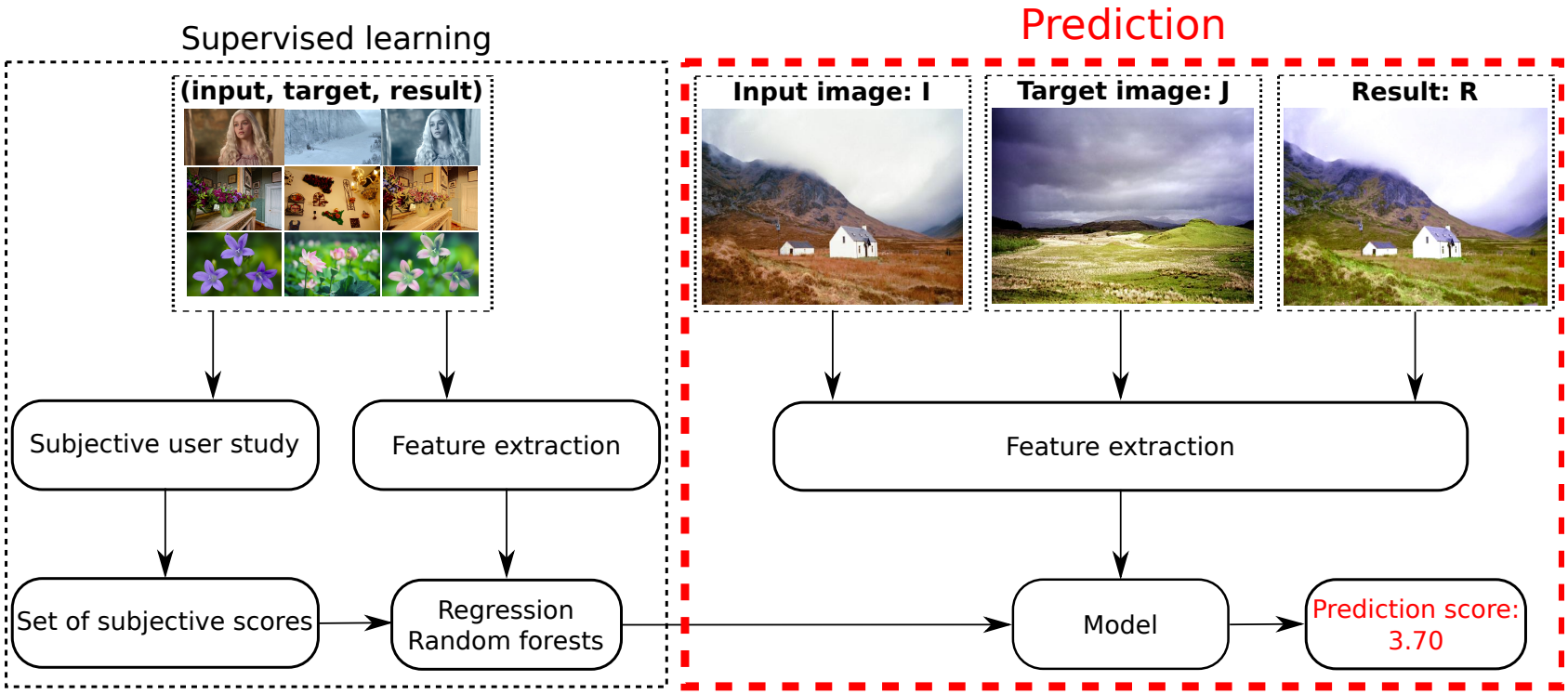
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

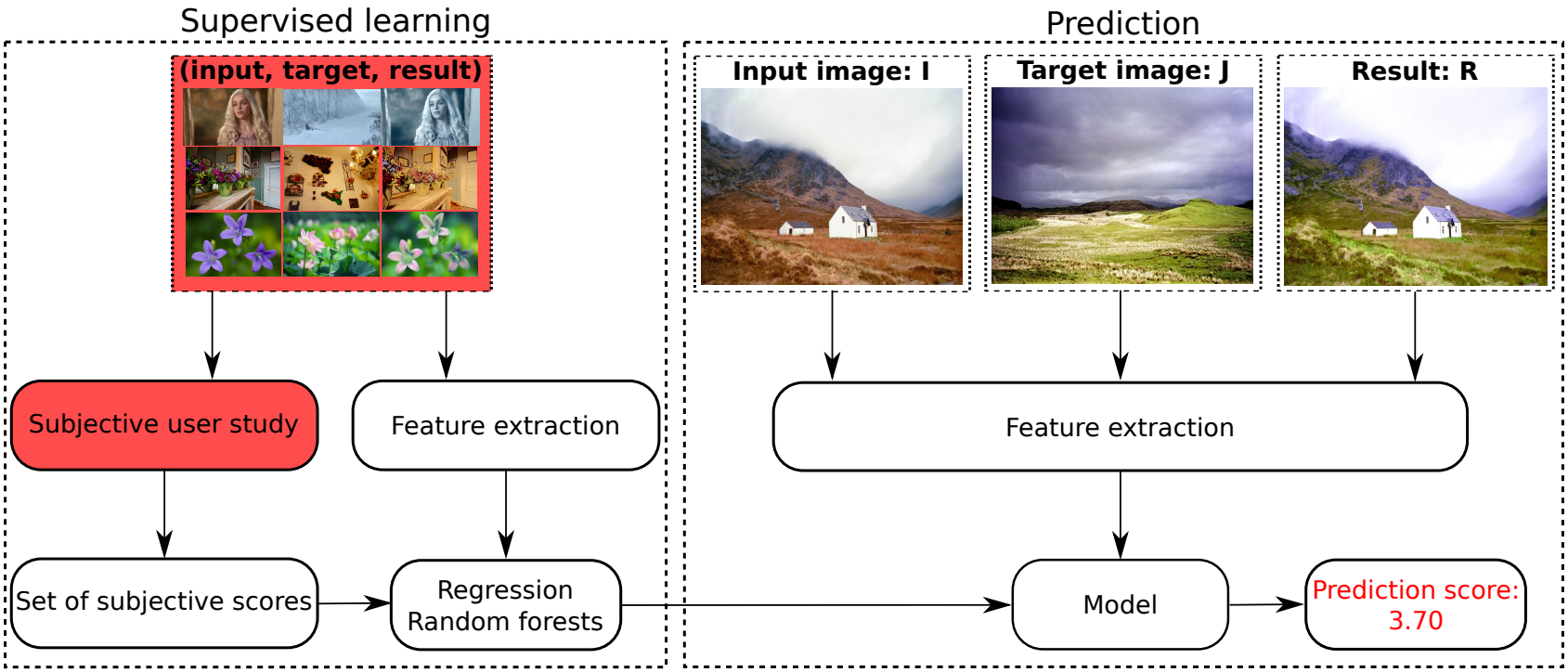
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

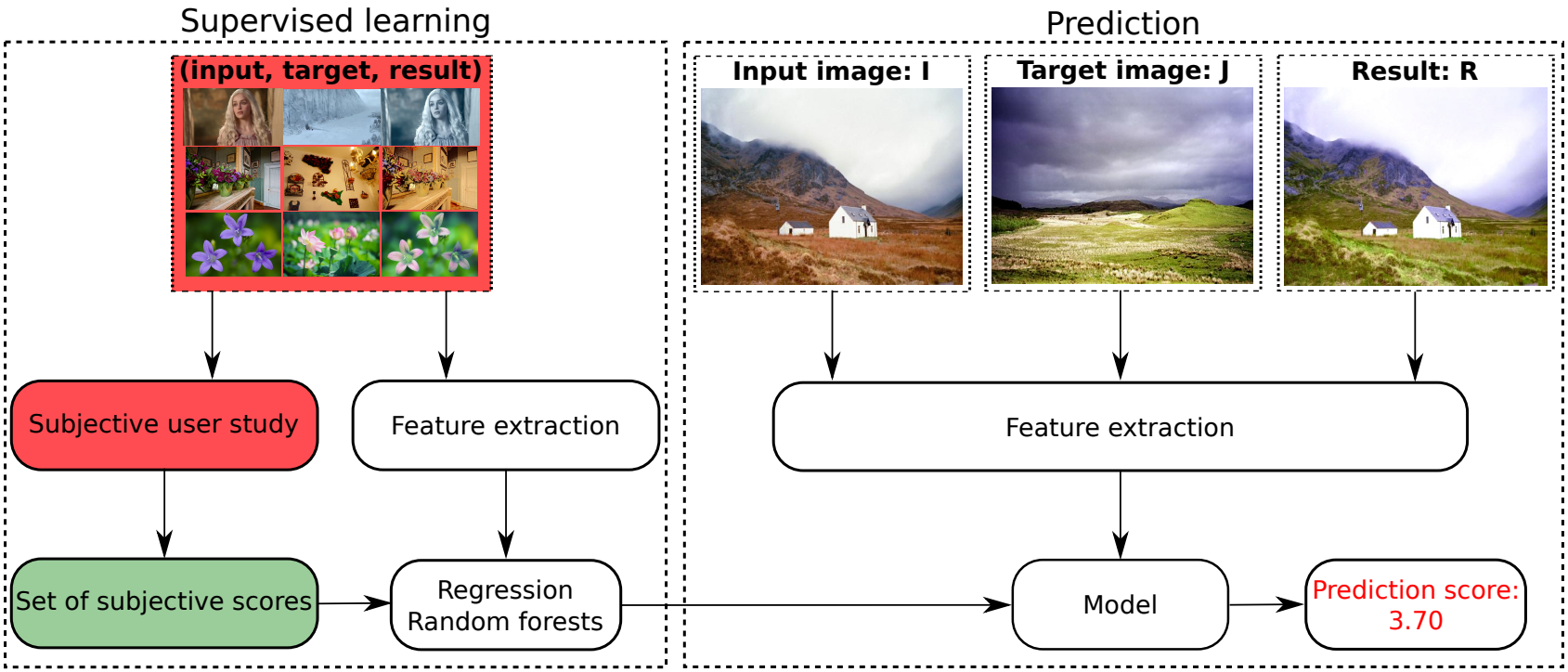
Flowchart


Supervised learning
Prediction
(input, target, result)
Input image: I
Target image: J
Result: R
Subjective user study
Feature extraction
Feature extraction
Set of subjective scores
Regression
Random forests
Model
Prediction score:
3.70

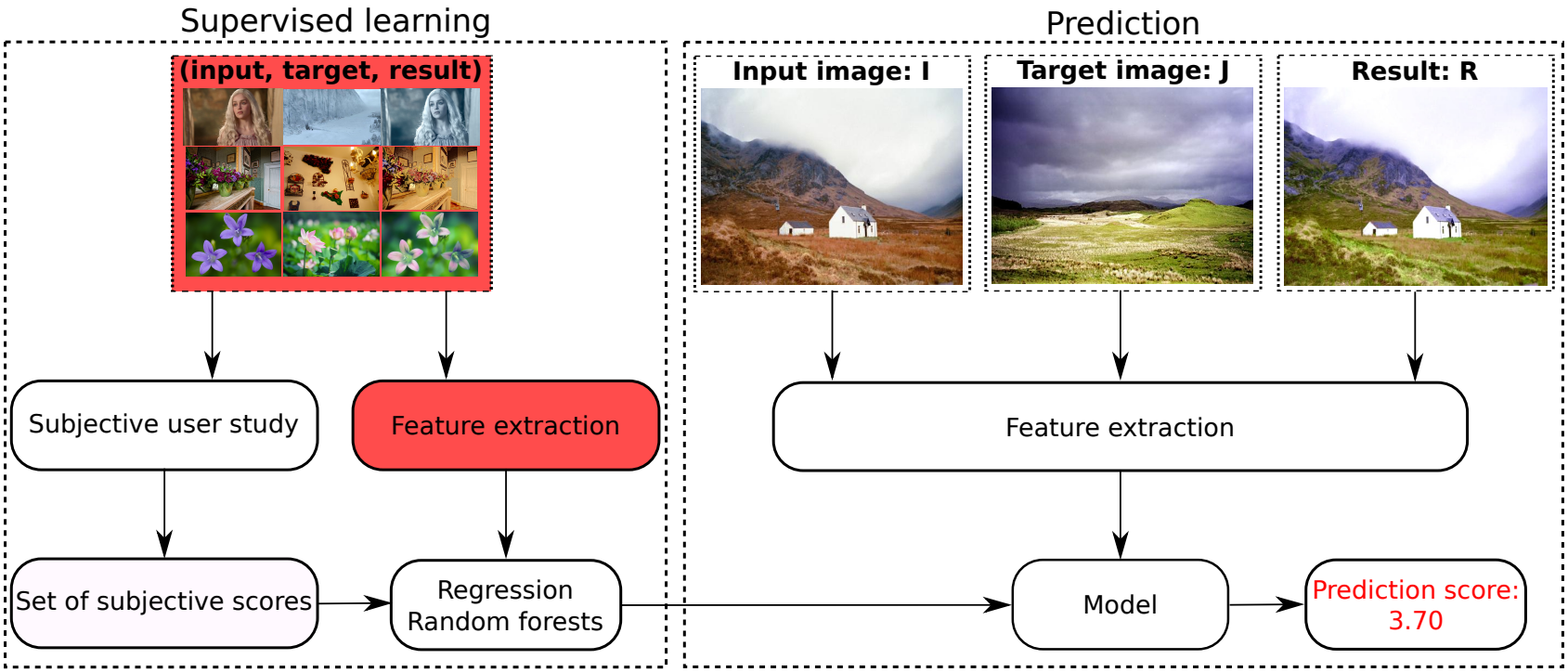
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

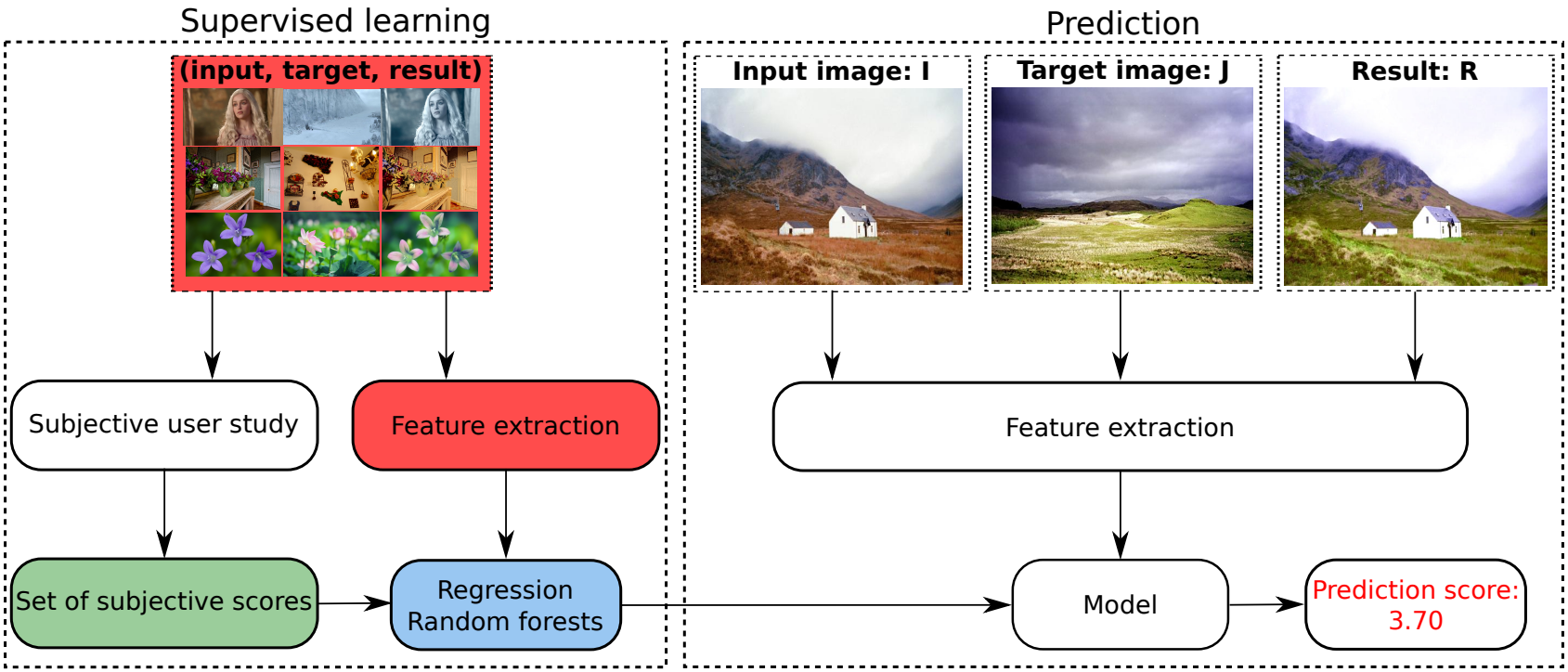
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

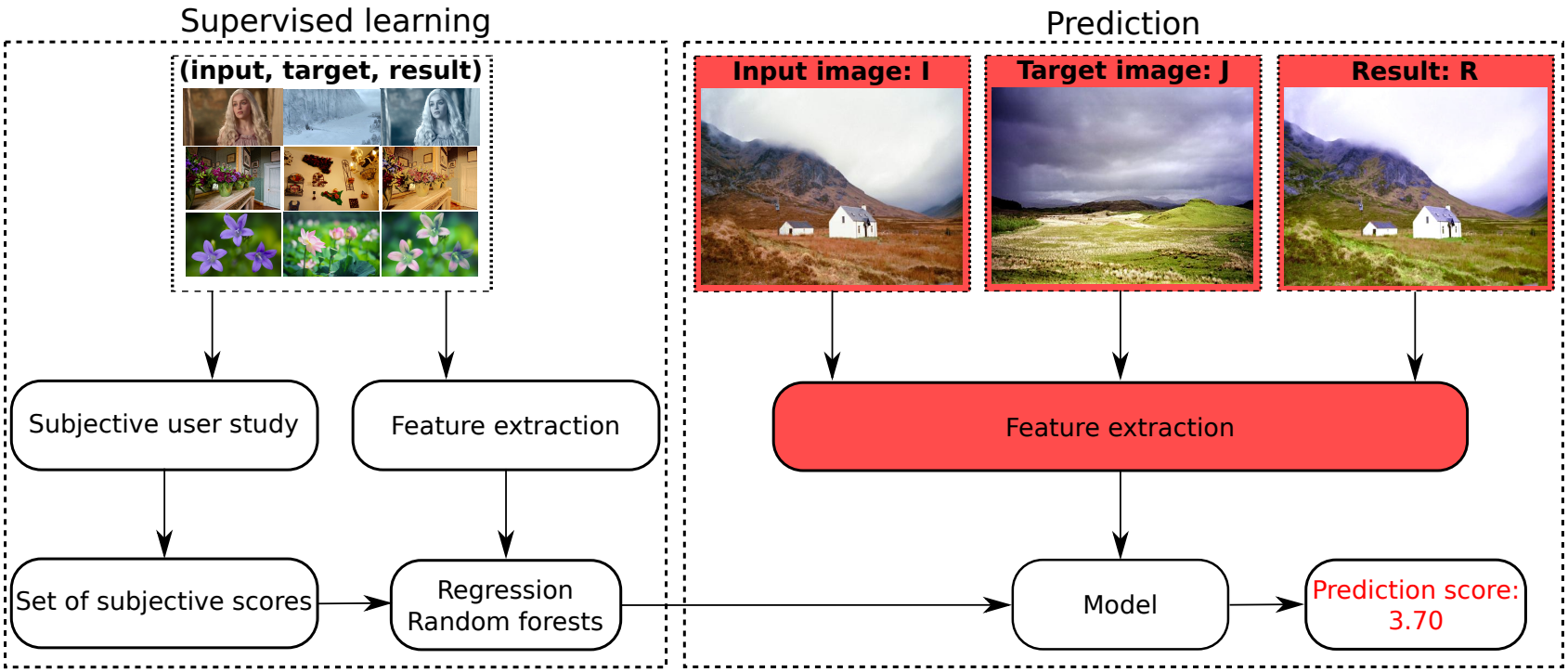
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

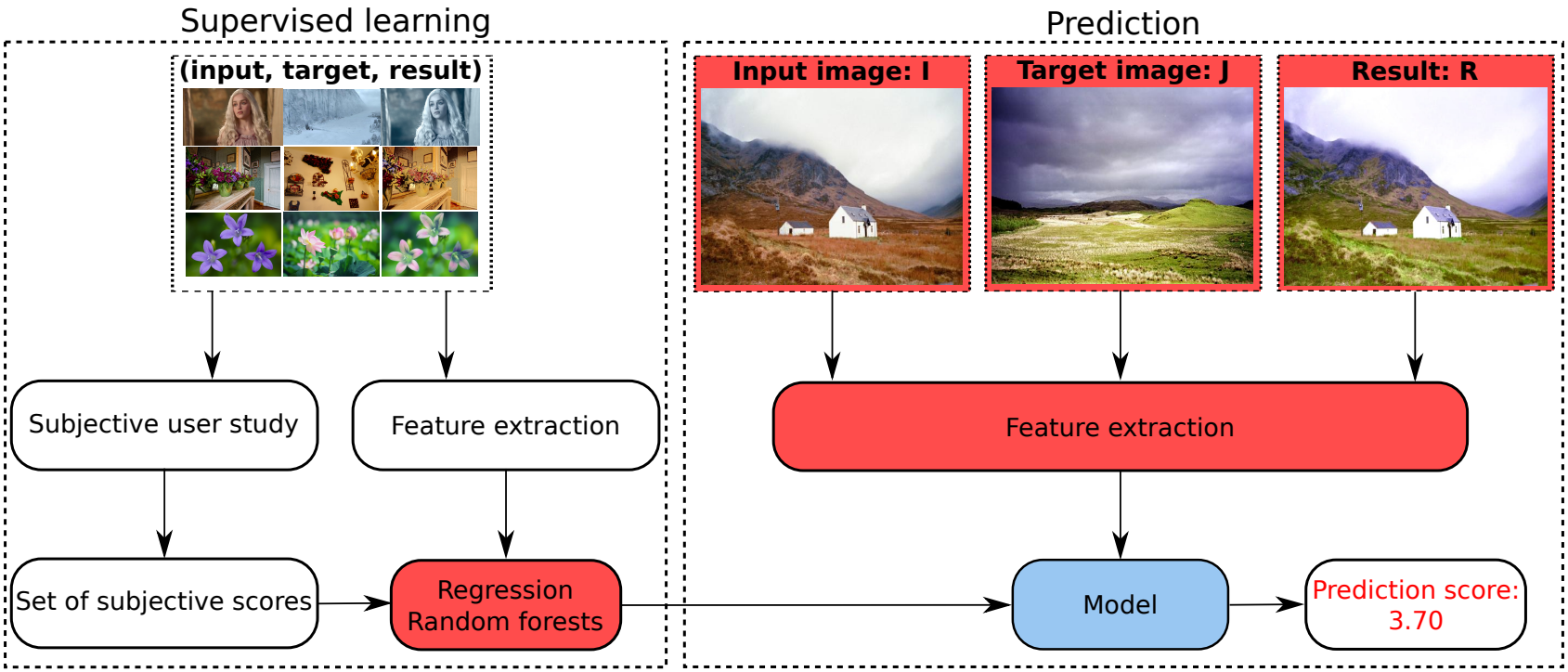
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

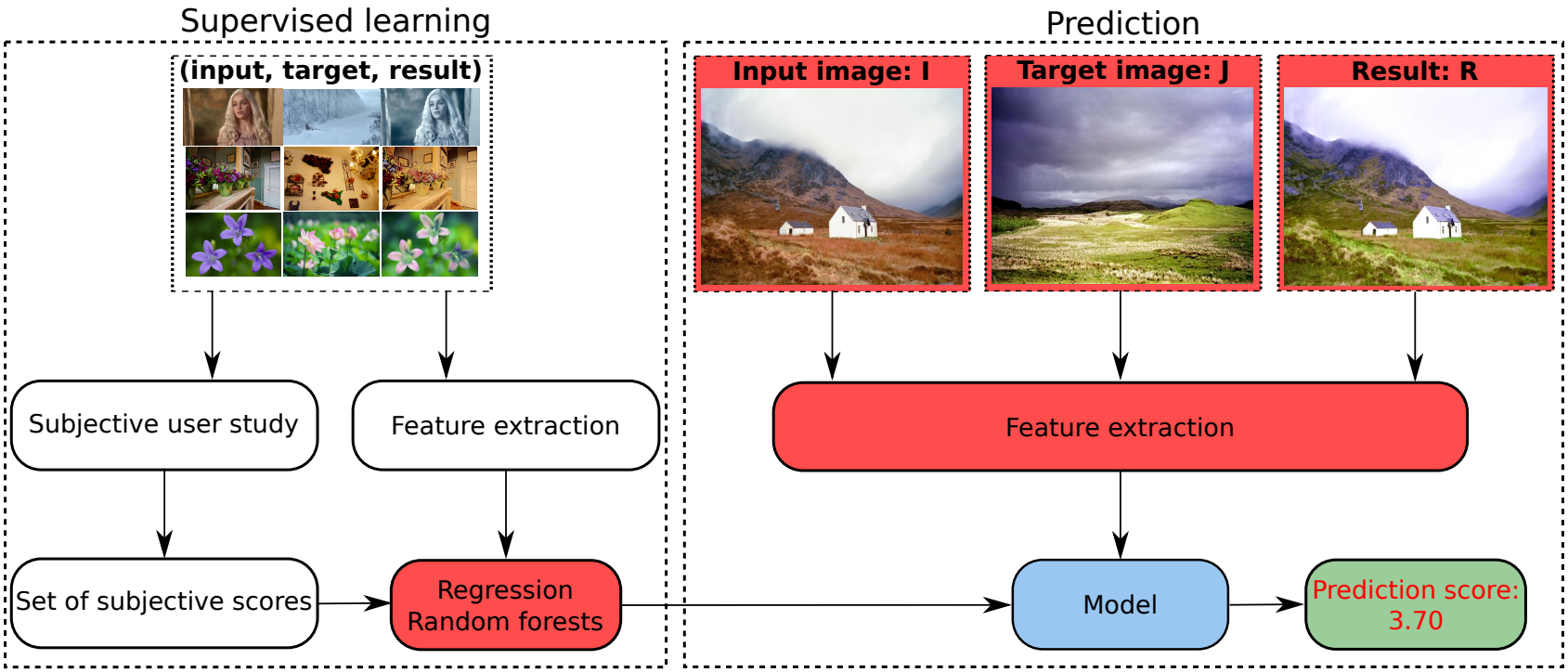
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

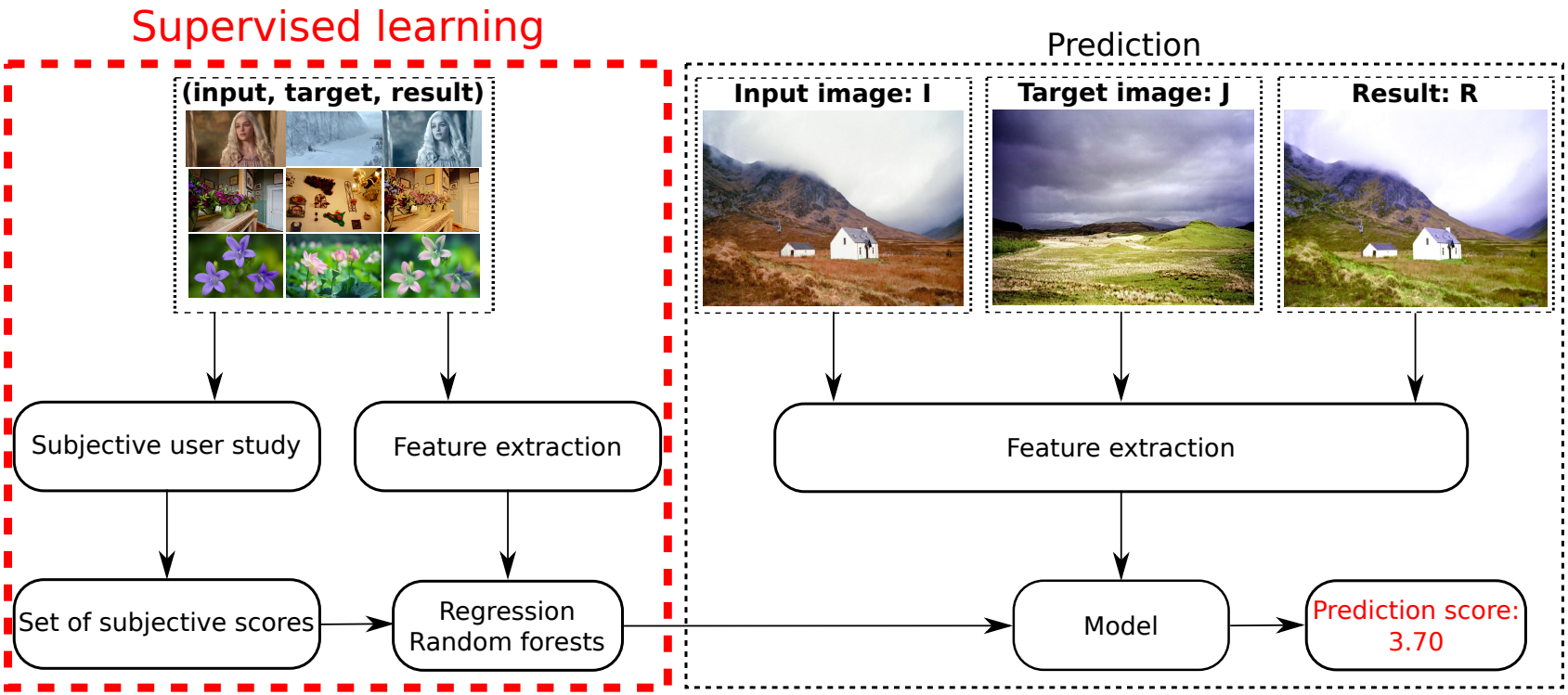
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

Outline

User study

Protocol

User study

Protocol

- 6 color transfer methods;

User study

Protocol

- 6 color transfer methods;
- 120 different results from color transfer methods;

User study

Protocol

- 6 color transfer methods;
- 120 different results from color transfer methods;
- Triplets of *(input, result, target)*;

User study

Protocol

- 6 color transfer methods;
- 120 different results from color transfer methods;
- Triplets of *(input, result, target)*;
- Each result - shown 5 seconds later than the input and target images;

User study

Protocol

- 6 color transfer methods;
- 120 different results from color transfer methods;
- Triplets of *(input, result, target)*;
- Each result - shown 5 seconds later than the input and target images;
- Two evaluation tasks:

User study

Protocol

- 6 color transfer methods;
- 120 different results from color transfer methods;
- Triplets of *(input, result, target)*;
- Each result - shown 5 seconds later than the input and target images;
- Two evaluation tasks:
  1. Evaluating the match in color between the result and the target image;

User study

Protocol

- 6 color transfer methods;
- 120 different results from color transfer methods;
- Triplets of *(input, result, target)*;
- Each result - shown 5 seconds later than the input and target images;
- Two evaluation tasks:
  1. Evaluating the match in color between the result and the target image;
  2. Users judgment on the aesthetic quality of the result.

User study

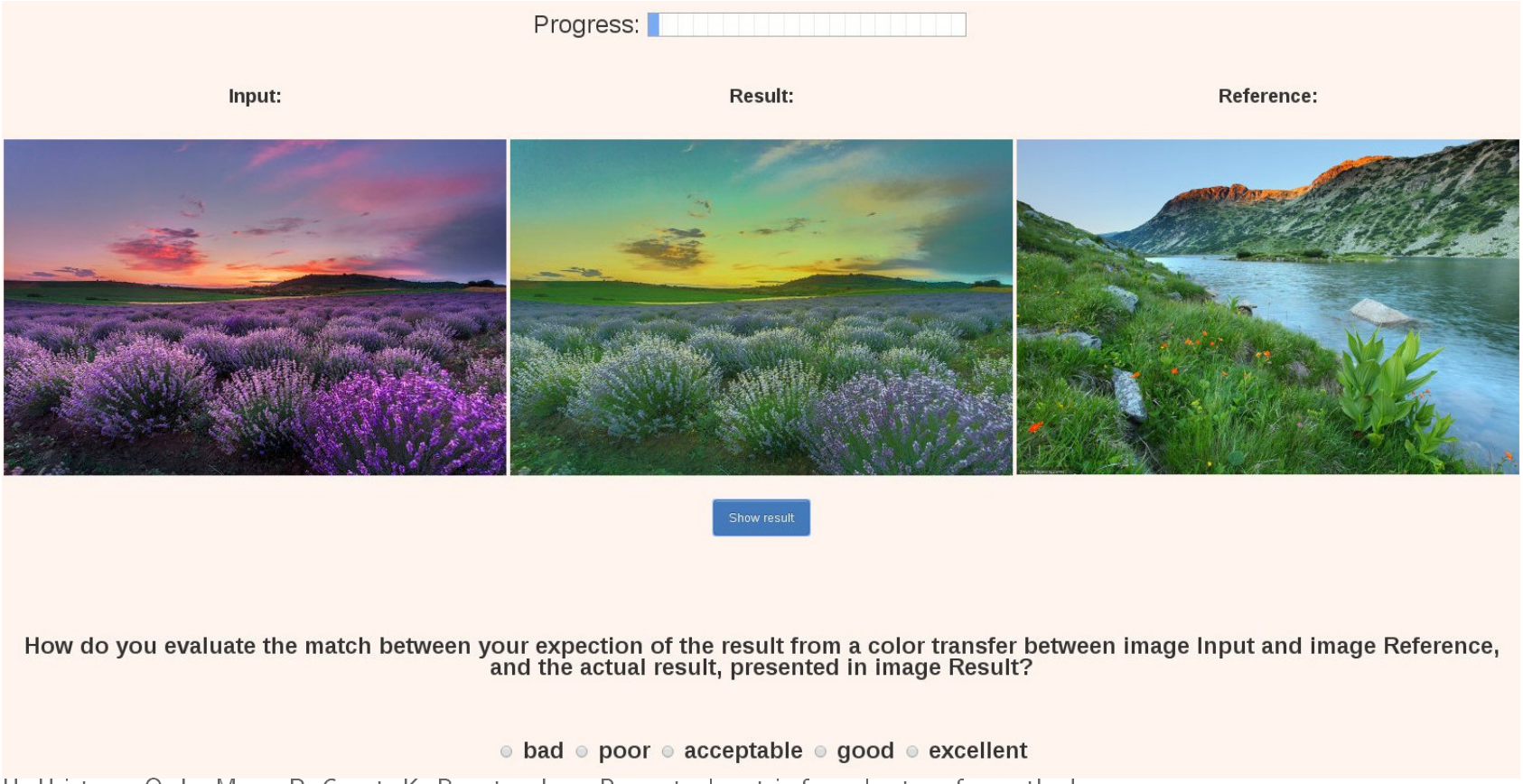
Protocol

- 6 color transfer methods;
- 120 different results from color transfer methods;
- Triplets of *(input, result, target)*;
- Each result - shown 5 seconds later than the input and target images;
- Two evaluation tasks:
  1. Evaluating the match in color between the result and the target image;
  2. Users judgment on the aesthetic quality of the result.
- 5-point scale.

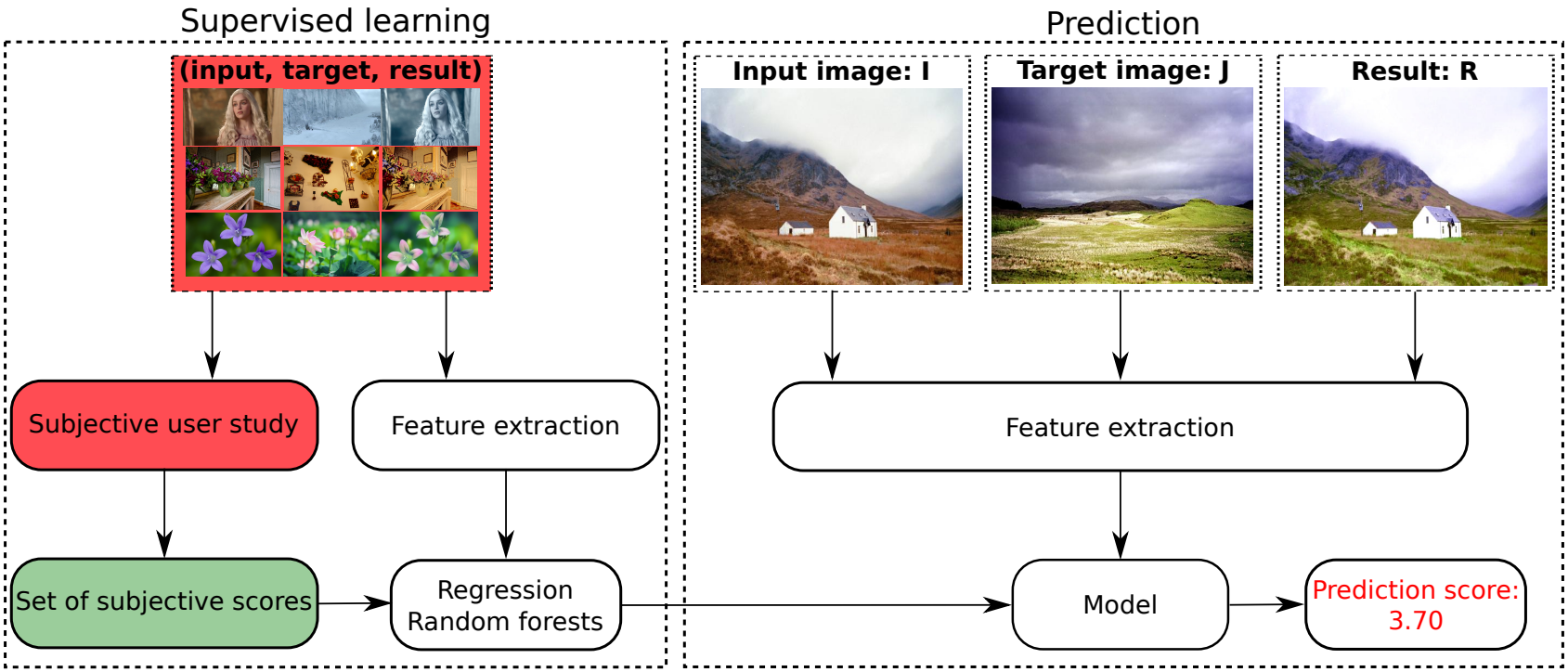
User study

Platform

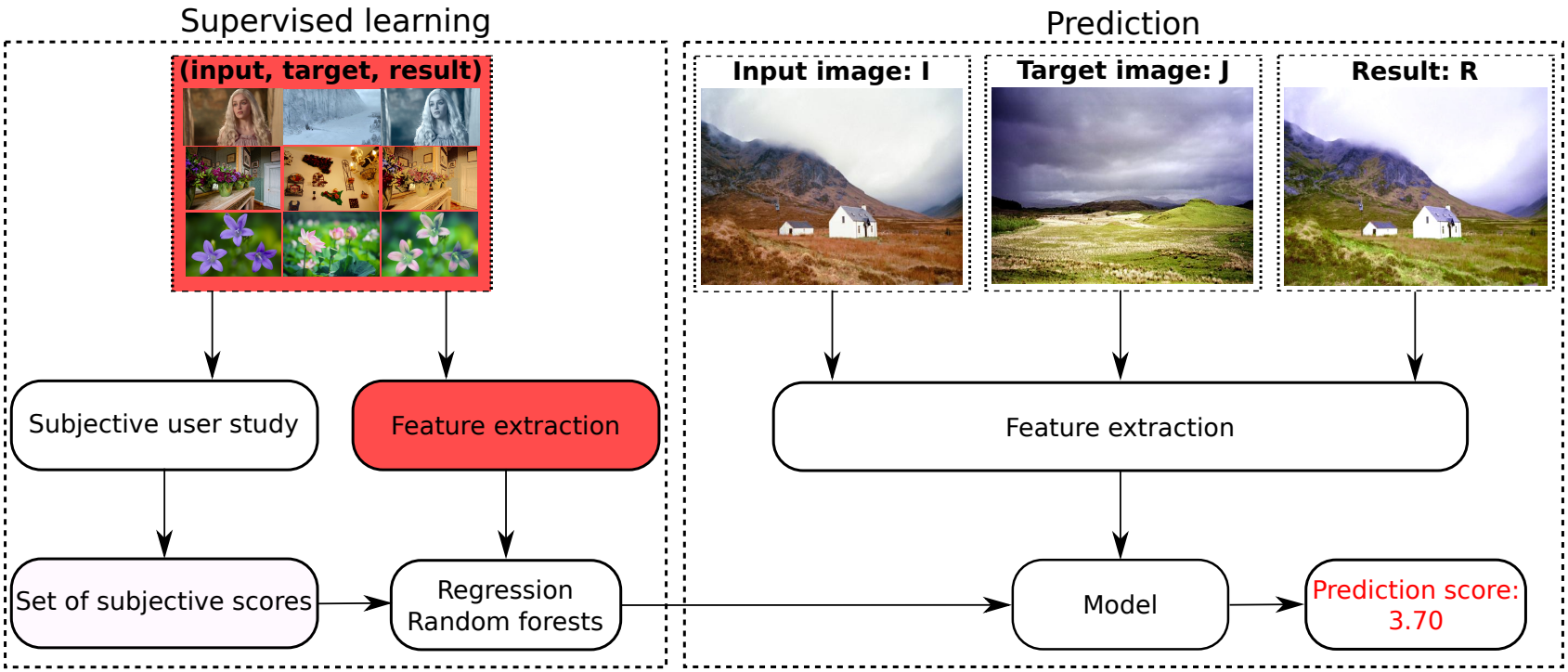
20 users evaluated the results of the color transfer methods individually using our specially designed *online platform*.



Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

Feature extraction

Objective features

Feature extraction

Objective features

1. SSIM(input,result);

Feature extraction

Objective features

1. SSIM(input,result);
2. Luminance similarity: $Bhatt(result_L, target_L)$;

Feature extraction

Objective features

1. SSIM(input,result);
2. Luminance similarity: $Bhatt(result_L, target_L)$;
3. Color similarity: $Bhatt(result_{ab}, target_{ab})$;

Feature extraction

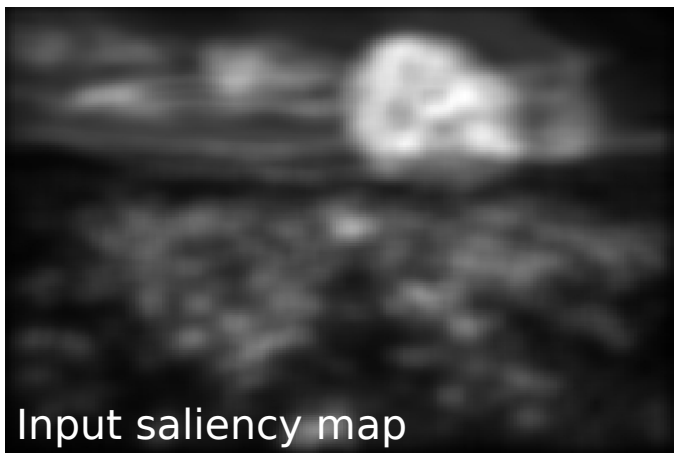
Objective features

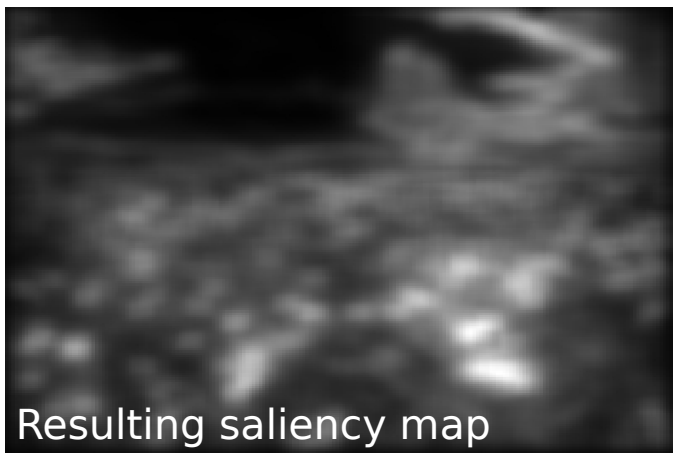
1. SSIM(input,result);
2. Luminance similarity: $Bhatt(result_L, target_L)$;
3. Color similarity: $Bhatt(result_{ab}, target_{ab})$;
4. Saliency similarity: $SalSim(input, result)$ [LMB13].

Feature extraction

Objective features

1. SSIM(input,result);
2. Luminance similarity: $Bhatt(result_L, target_L)$;
3. Color similarity: $Bhatt(result_{ab}, target_{ab})$;
4. Saliency similarity: $SalSim(input, result)$ [LMB13].





Feature extraction

Objective features

Feature extraction

Objective features

5. Histogram similarity of each of 5 color appearance attributes [FJ04], *i.e.* brightness, lightness, chroma, colorfulness, saturation: $Bhatt(result_{attr}, target_{attr})$.

Feature extraction

Objective features

5. Histogram similarity of each of 5 color appearance attributes [FJ04], *i.e.* brightness, lightness, chroma, colorfulness, saturation: $Bhatt(result_{attr}, target_{attr})$.

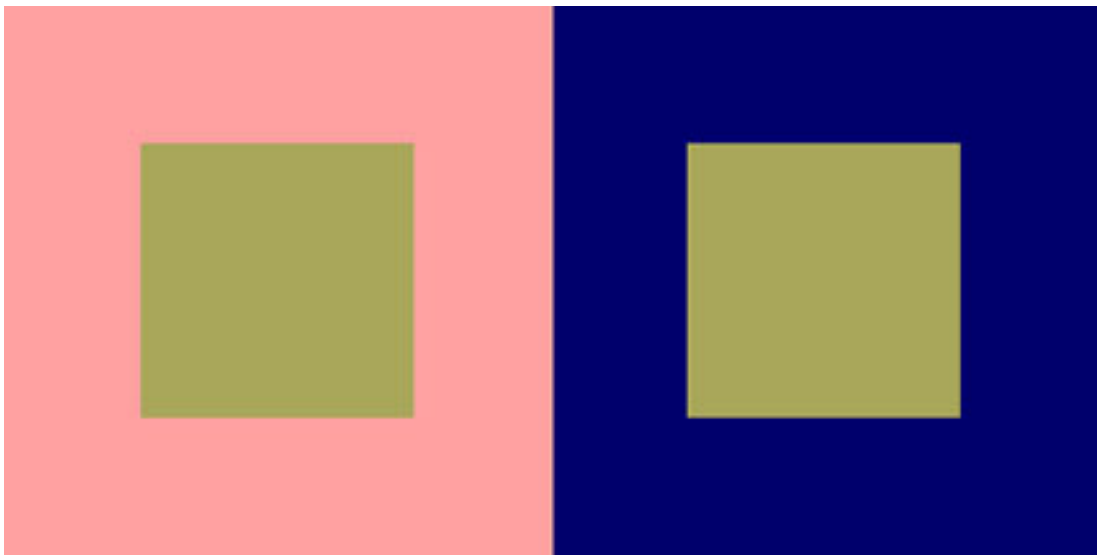
The color appearance attributes describe the perceptual aspects of color and account for various color appearance phenomena.

Feature extraction

Objective features

5. Histogram similarity of each of 5 color appearance attributes [FJ04], *i.e.* brightness, lightness, chroma, colorfulness, saturation: $Bhatt(result_{attr}, target_{attr})$.

The color appearance attributes describe the perceptual aspects of color and account for various color appearance phenomena.

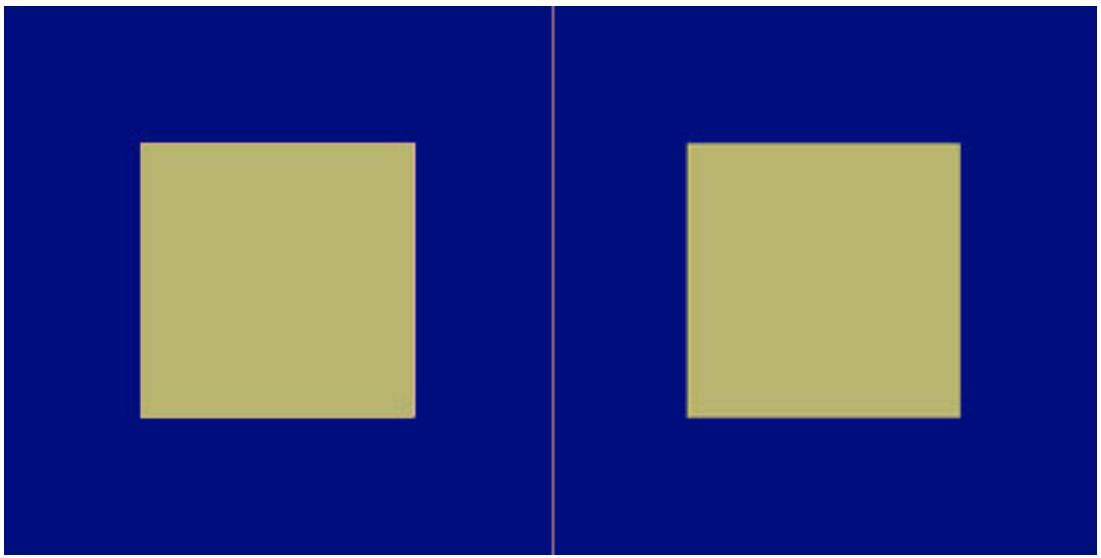


Feature extraction

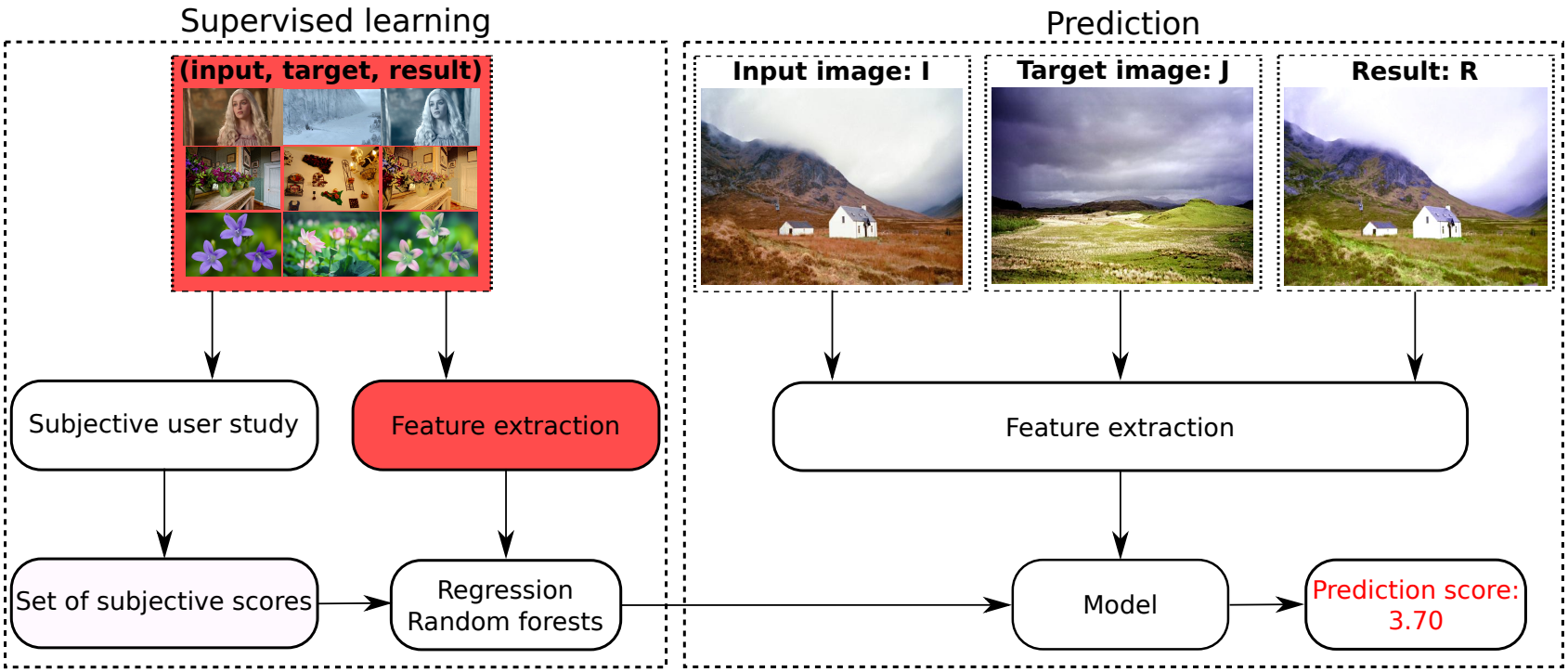
Objective features

5. Histogram similarity of each of 5 color appearance attributes [FJ04], *i.e.* brightness, lightness, chroma, colorfulness, saturation: $Bhatt(result_{attr}, target_{attr})$.

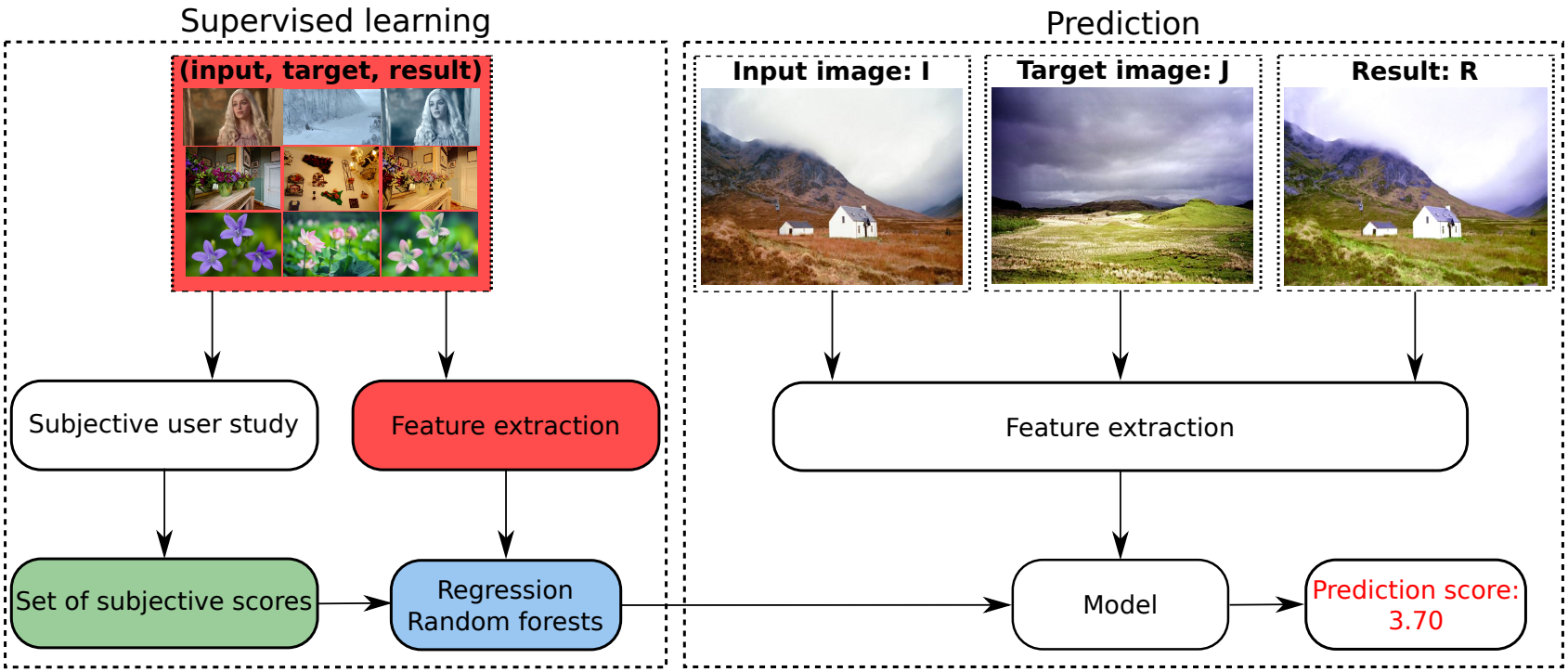
The color appearance attributes describe the perceptual aspects of color and account for various color appearance phenomena.



Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

Perceptual model

Main goal

How can we describe the perceptual scores from the user study using the set of extracted image features?

Perceptual model

Main goal

How can we describe the perceptual scores from the user study using the set of extracted image features?

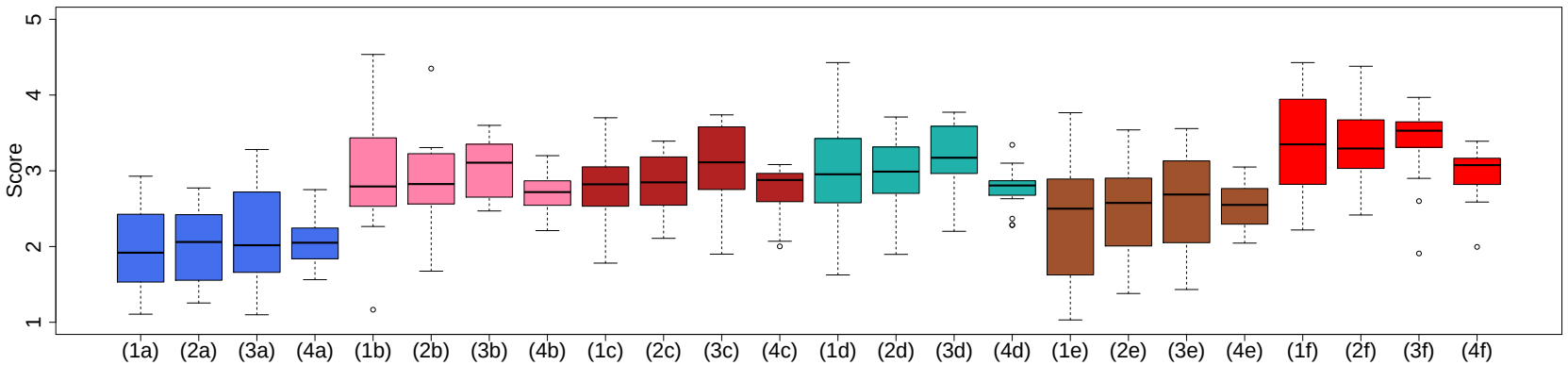
Solutions

- Linear regression;
- Non-linear regression;
- Random forests.

Perceptual model

Data from user study explained

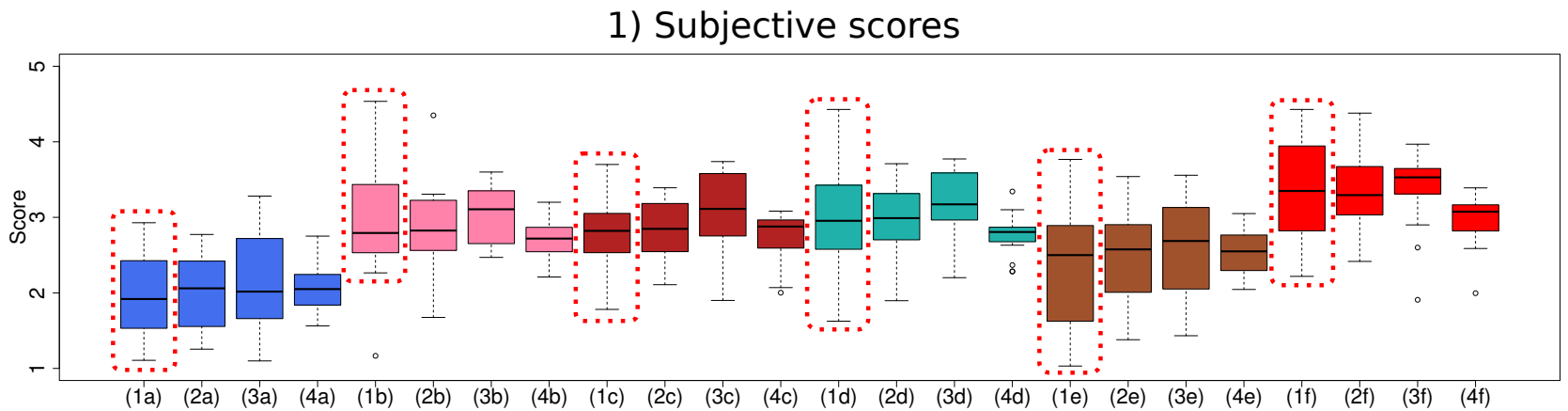
- **Box-and-Whisker plots for the score distributions;**
- The model using random forests most accurately describes the subjective scores;
- *[Hristova et al., 2015]* outperforms the other 4 methods.



Perceptual model

Data from user study explained

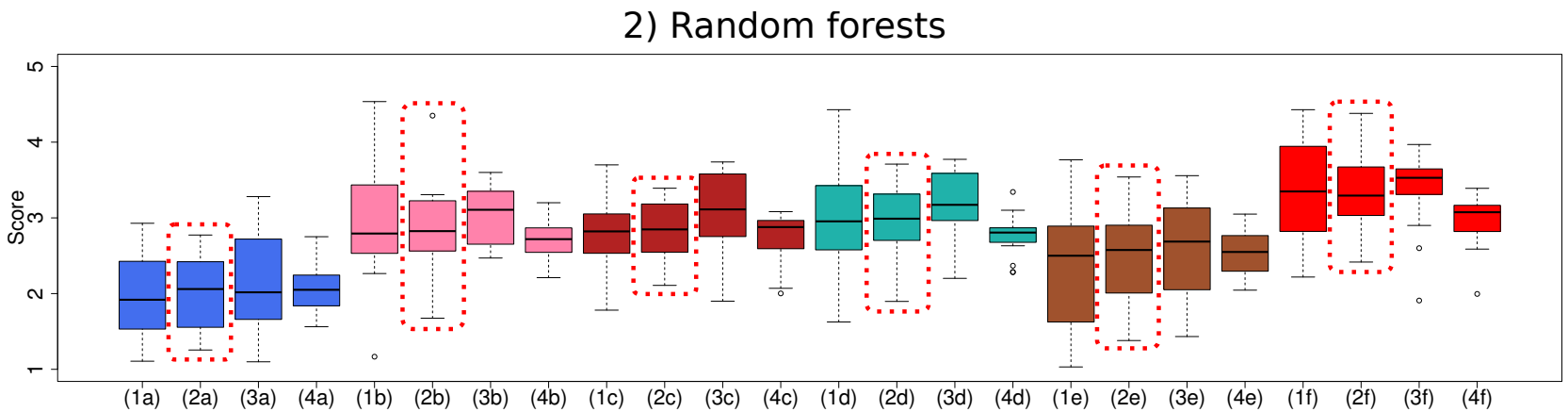
- Box-and-Whisker plots for the score distributions;
- The model using random forests most accurately describes the subjective scores;
- *[Hristova et al., 2015]* outperforms the other 4 methods.



Perceptual model

Data from user study explained

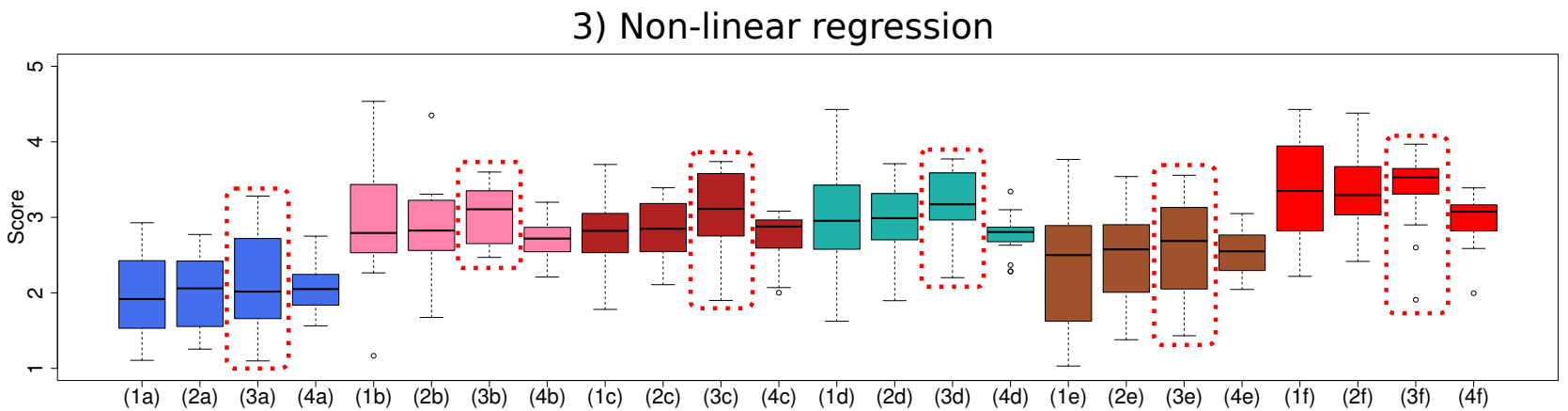
- Box-and-Whisker plots for the score distributions;
- The model using random forests most accurately describes the subjective scores;
- *[Hristova et al., 2015]* outperforms the other 4 methods.



Perceptual model

Data from user study explained

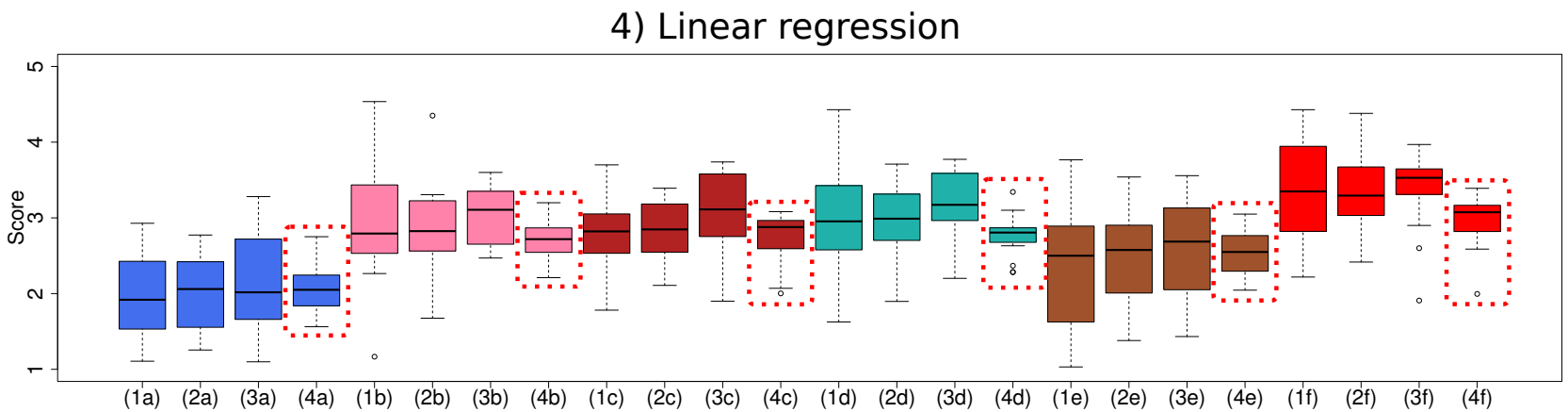
- Box-and-Whisker plots for the score distributions;
- The model using random forests most accurately describes the subjective scores;
- *[Hristova et al., 2015]* outperforms the other 4 methods.



Perceptual model

Data from user study explained

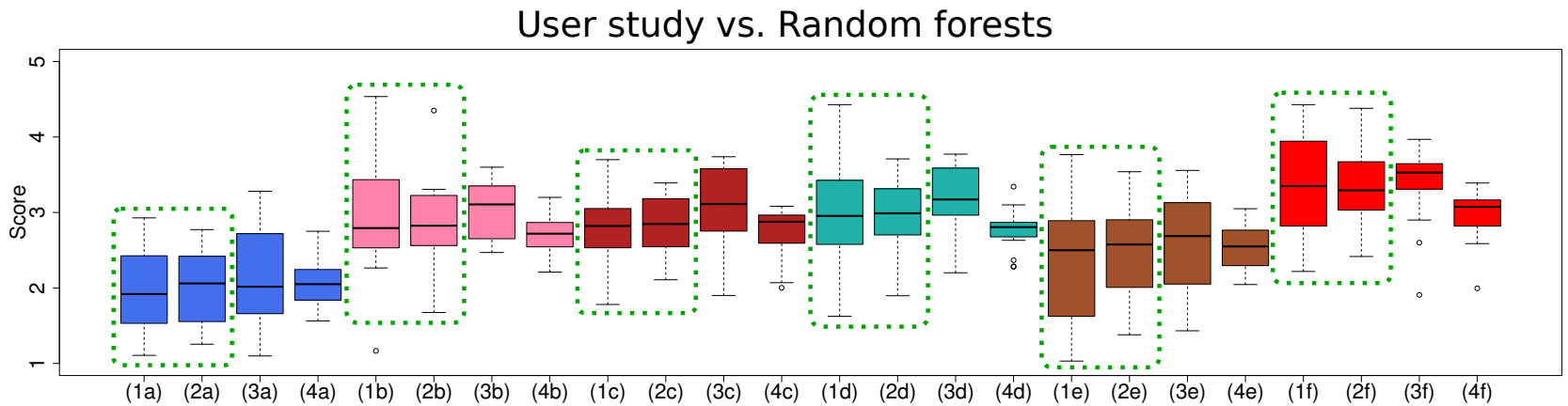
- Box-and-Whisker plots for the score distributions;
- The model using random forests most accurately describes the subjective scores;
- *[Hristova et al., 2015]* outperforms the other 4 methods.



Perceptual model

Data from user study explained

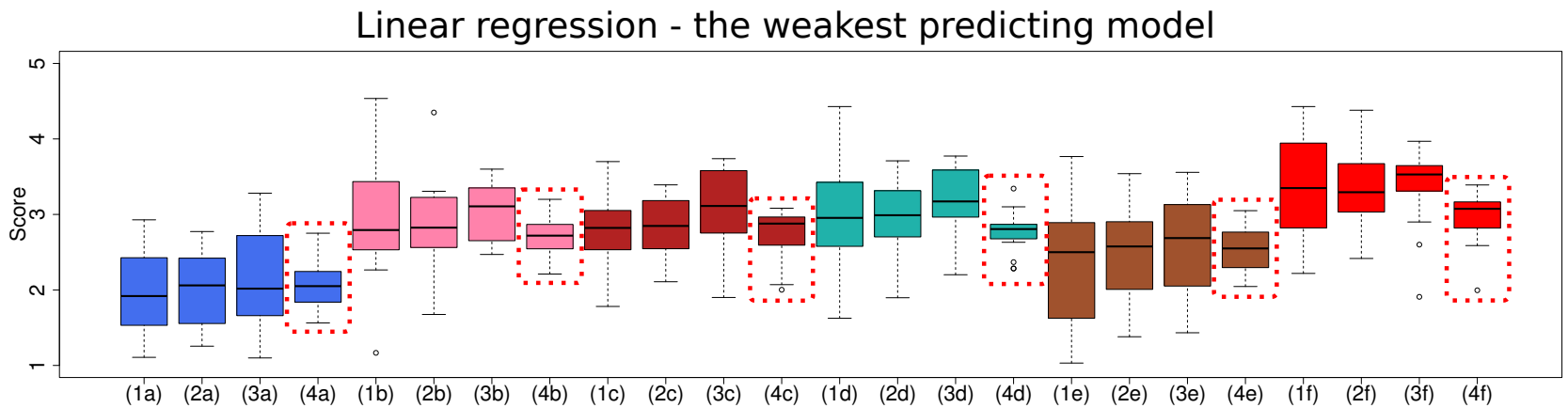
- Box-and-Whisker plots for the score distributions;
- **The model using random forests most accurately describes the subjective scores;**
- *[Hristova et al., 2015]* outperforms the other 4 methods.



Perceptual model

Data from user study explained

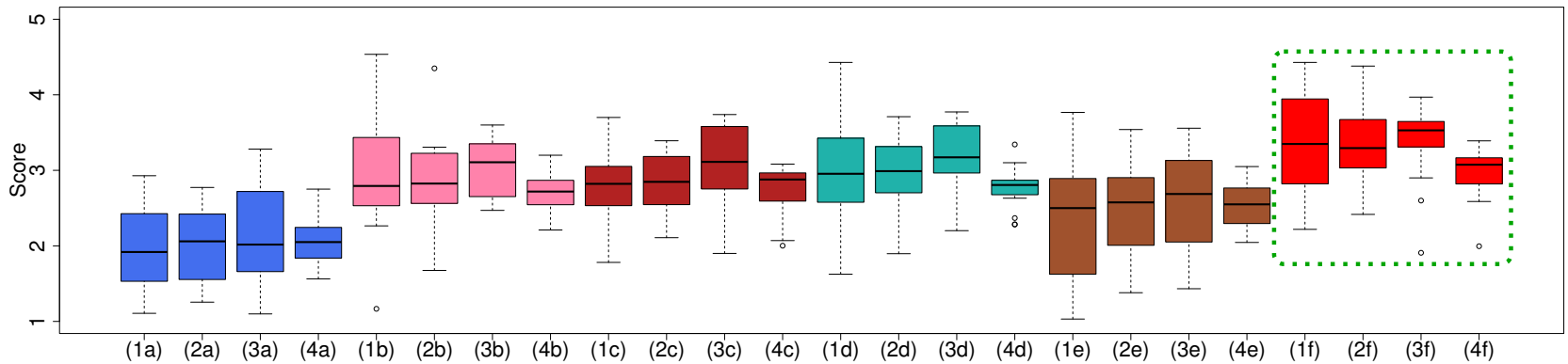
- Box-and-Whisker plots for the score distributions;
- The model using random forests most accurately describes the subjective scores;
- *[Hristova et al., 2015]* outperforms the other 4 methods.



Perceptual model

Data from user study explained

- Box-and-Whisker plots for the score distributions;
- The model using random forests most accurately describes the subjective scores;
- **[Hristova et al., 2015] outperforms the other 4 methods.**



Outline

Evaluation

Cross-validation

- K-fold cross-validation with 10 folds;

Evaluation

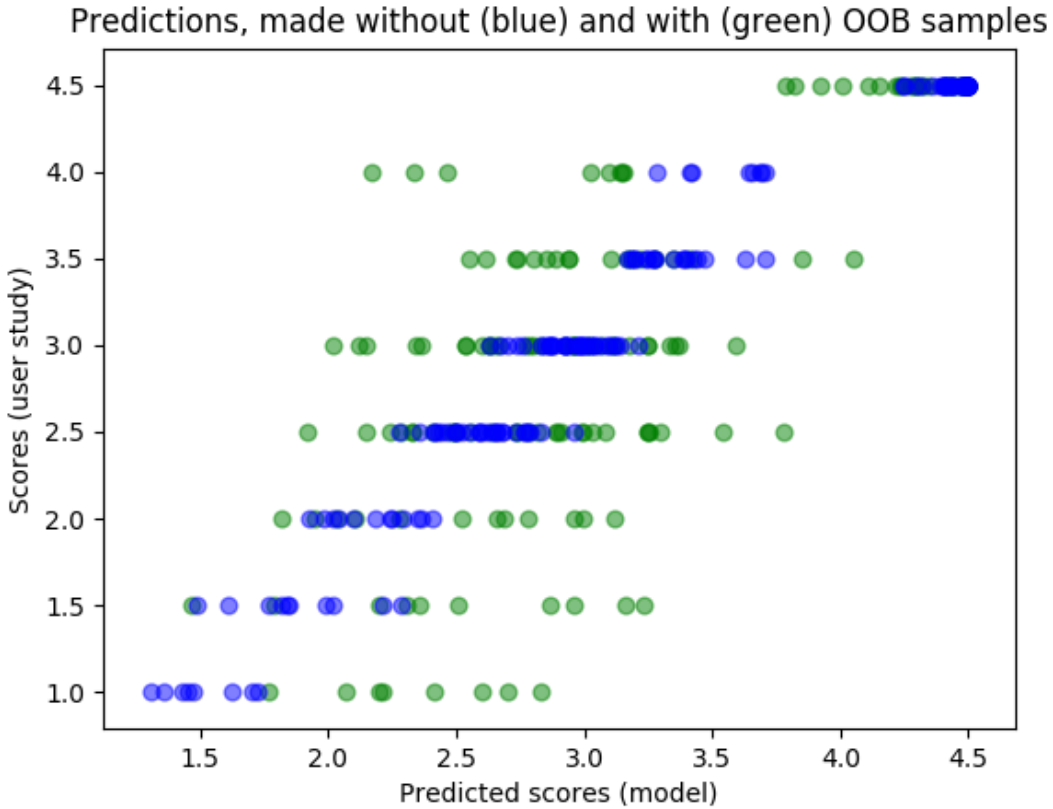
Cross-validation

- K-fold cross-validation with 10 folds;
- Correlation between the model and the user study;
- Mean square error (MSE).

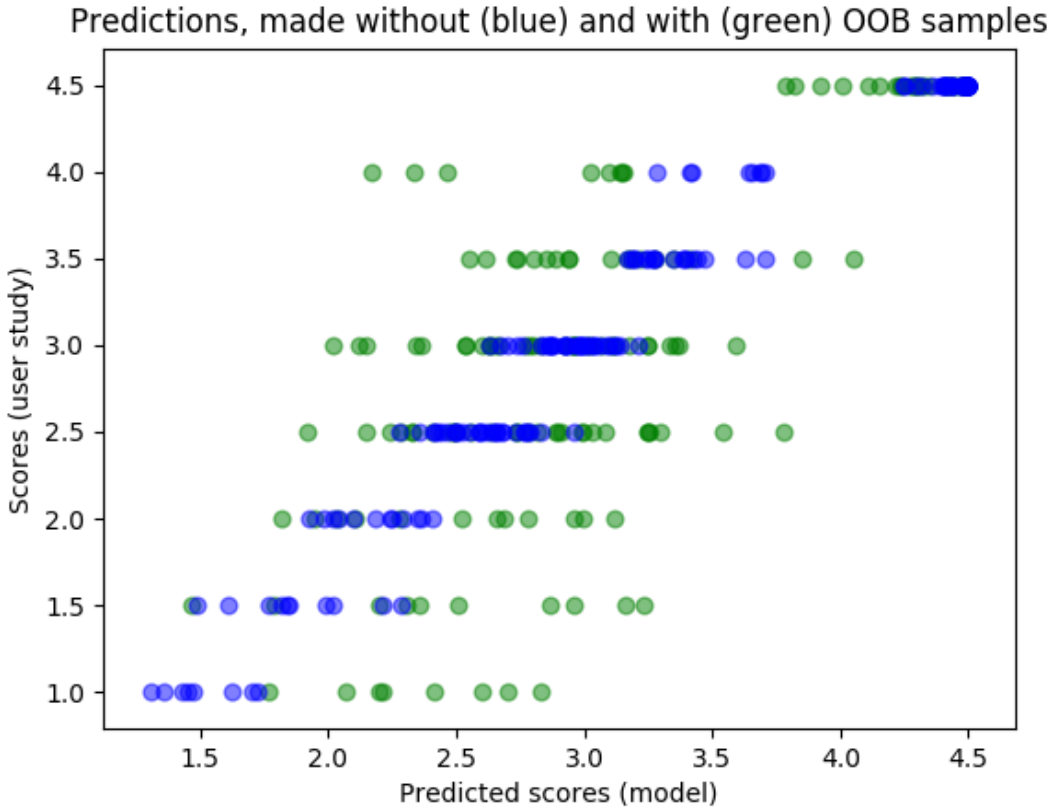
	Random forests	**Linear**	**Non-linear**
Corr	0.765 ± 0.133	0.567 ± 0.135	0.644 ± 0.157
MSE	0.472	0.808	0.889

Table: Correlation $\pm$ standard deviation and MSE over all 10 test sets in our cross-validation.

Subjective vs. Predicted scores

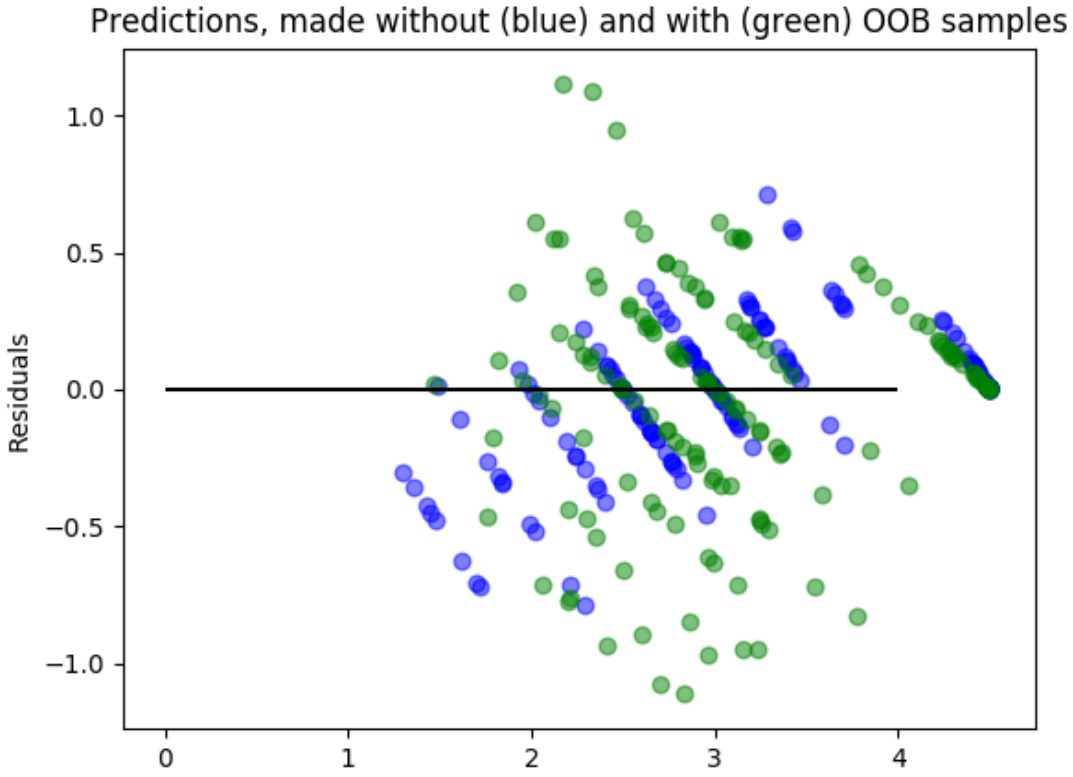


Subjective vs. Predicted scores

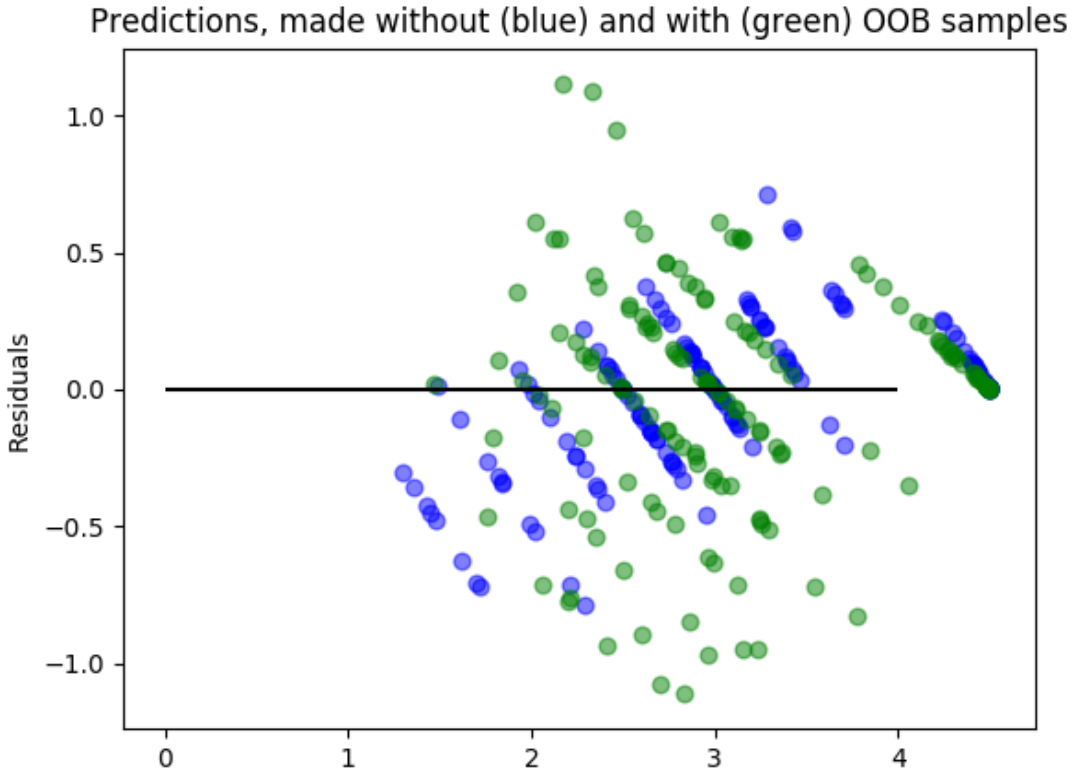


The plot shows the high correction between subjective scores and predicted scores (with and without OOB samples).

Residuals

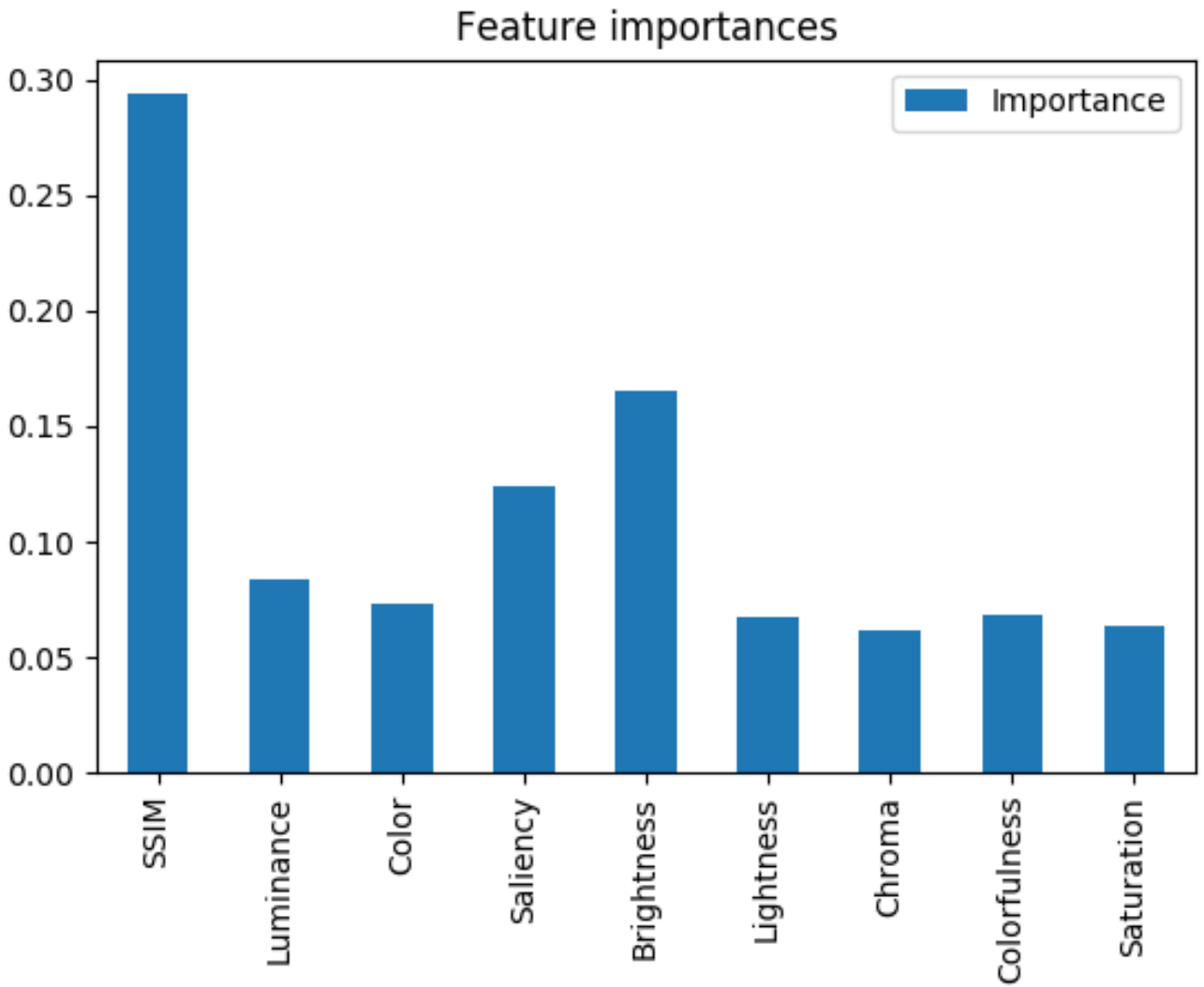

Predictions, made without (blue) and with (green) OOB samples
Residuals
1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4

Residuals

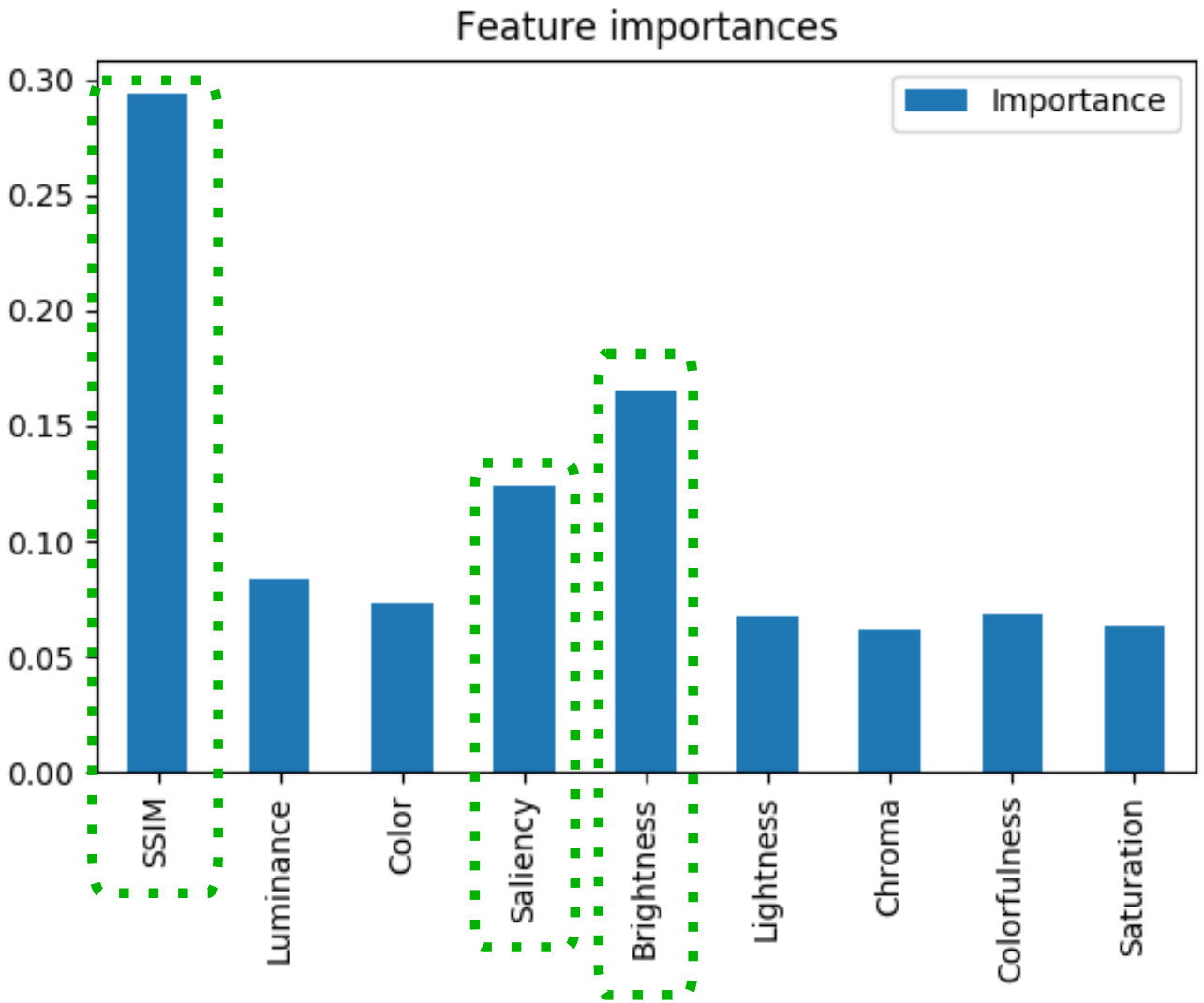


Residuals: in the interval [-1.5, 1.5] and symmetrical around the x-axis; random patterns.

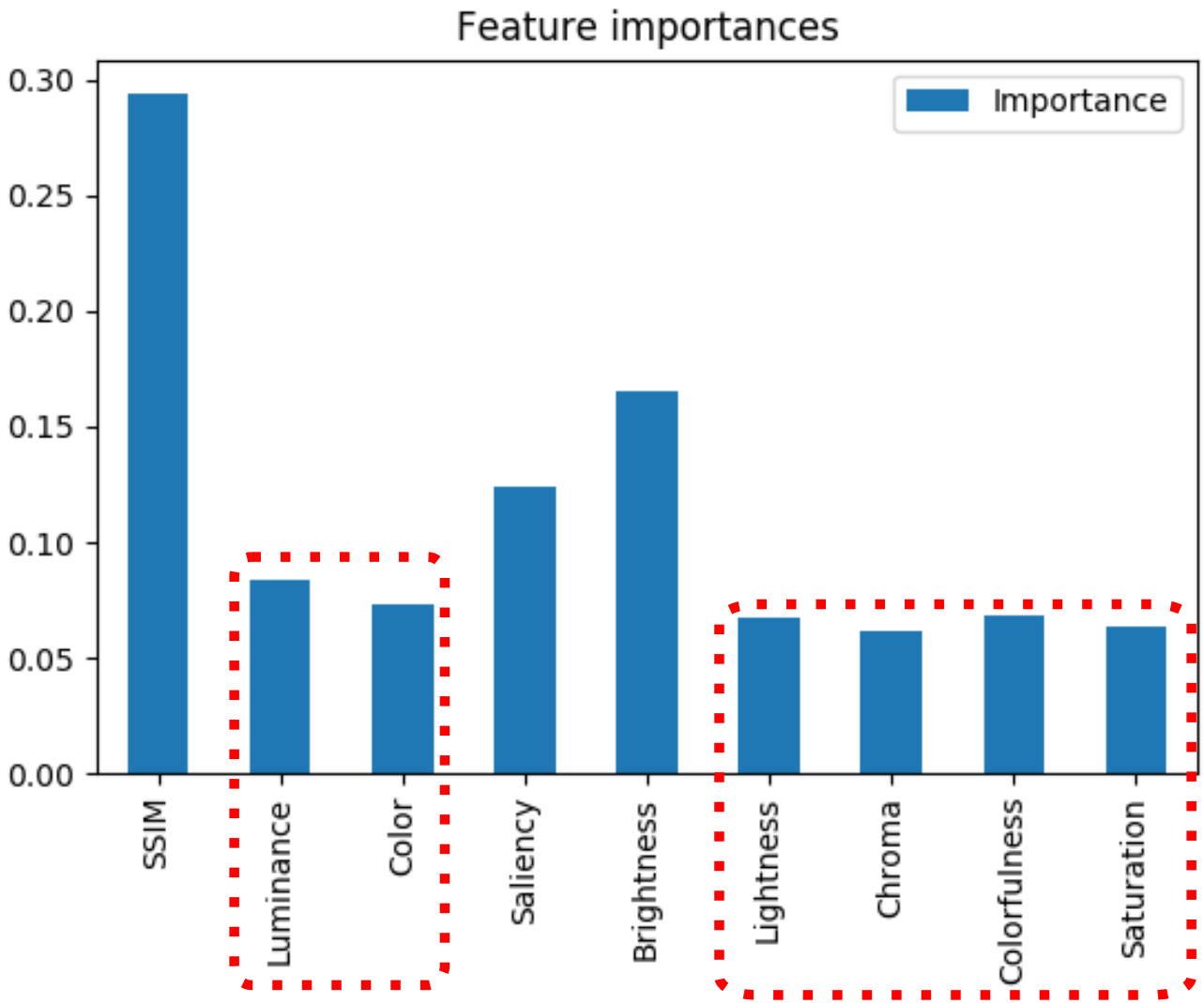
Feature importances



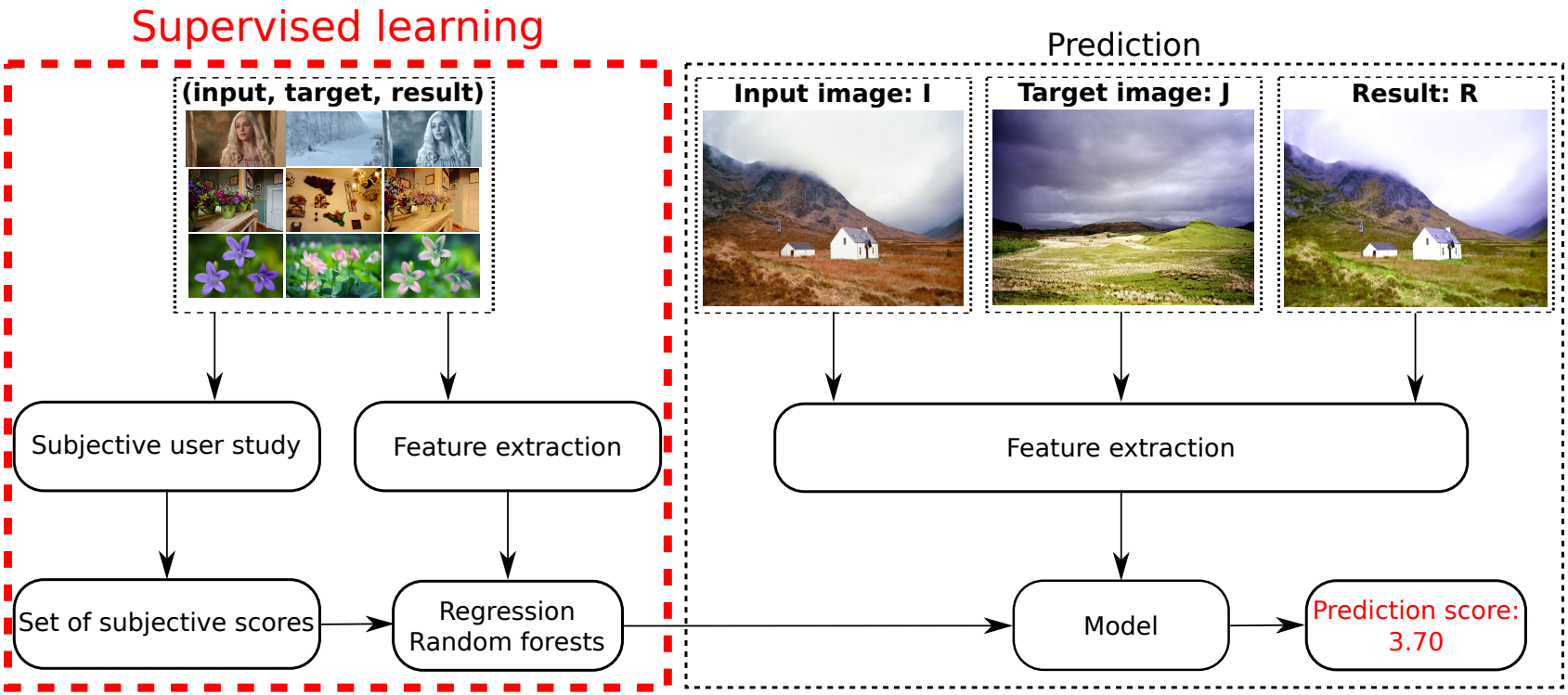
Feature importances



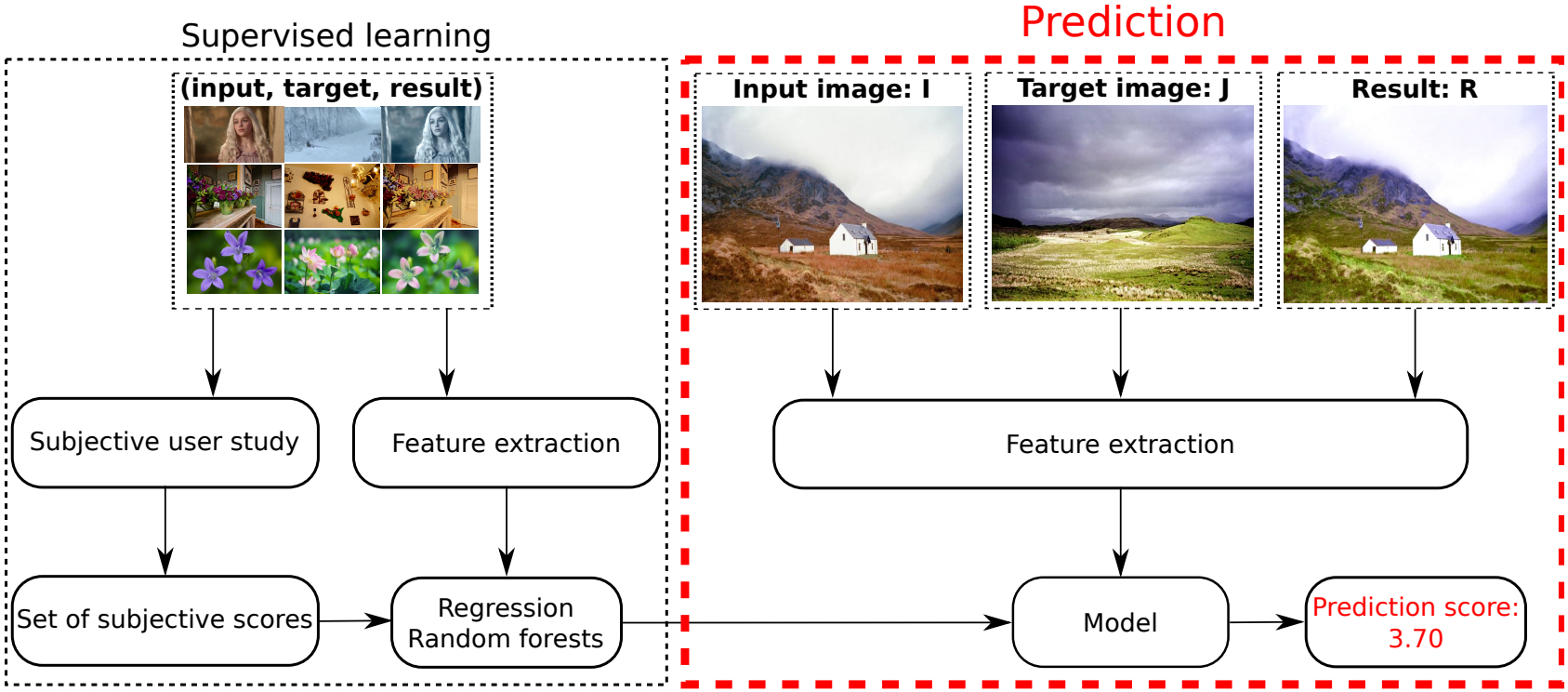
Feature importances



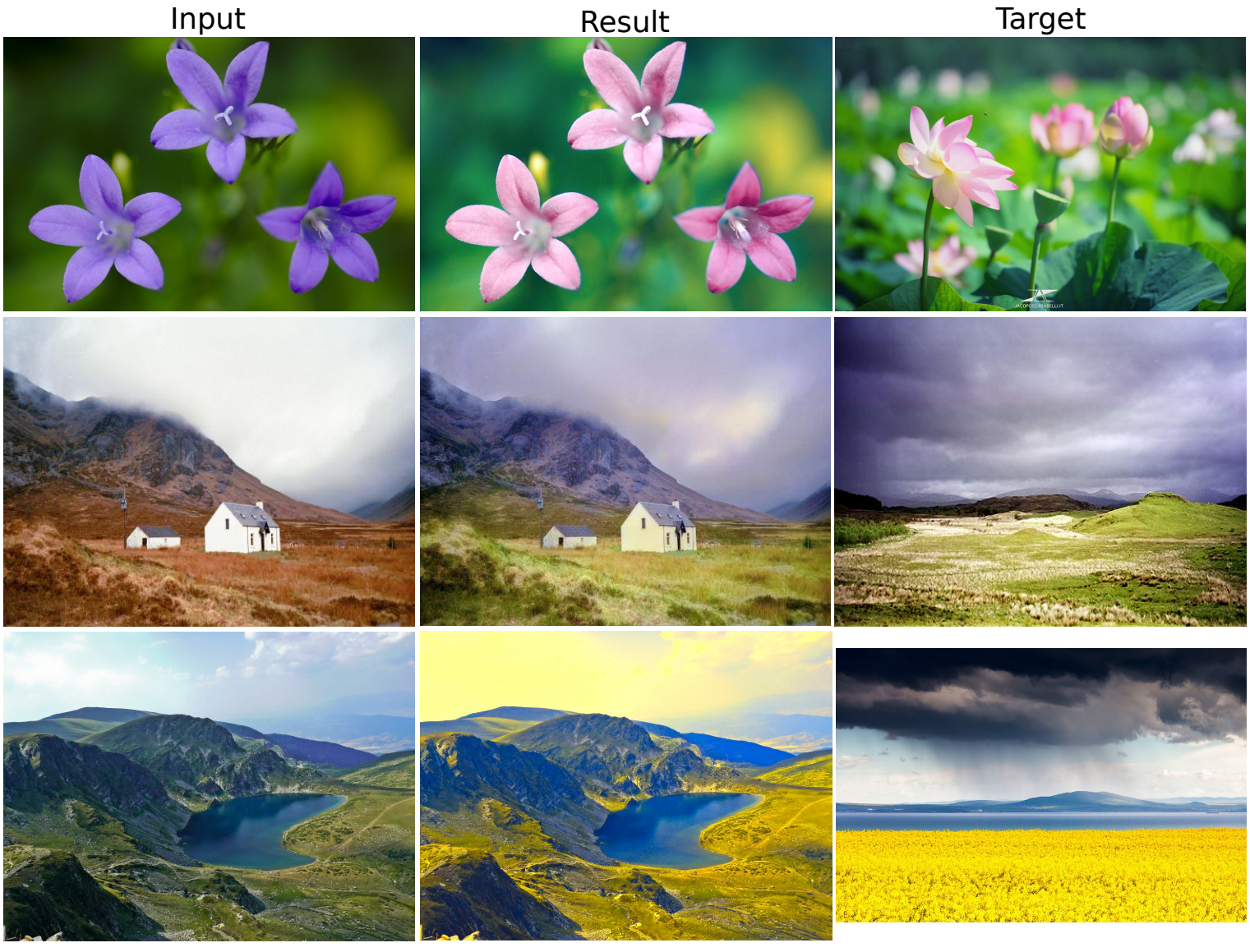
Flowchart


Supervised learning
(input, target, result)
Subjective user study
Feature extraction
Set of subjective scores
Regression
Random forests
Prediction
Input image: I
Target image: J
Result: R
Feature extraction
Model
Prediction score:
3.70

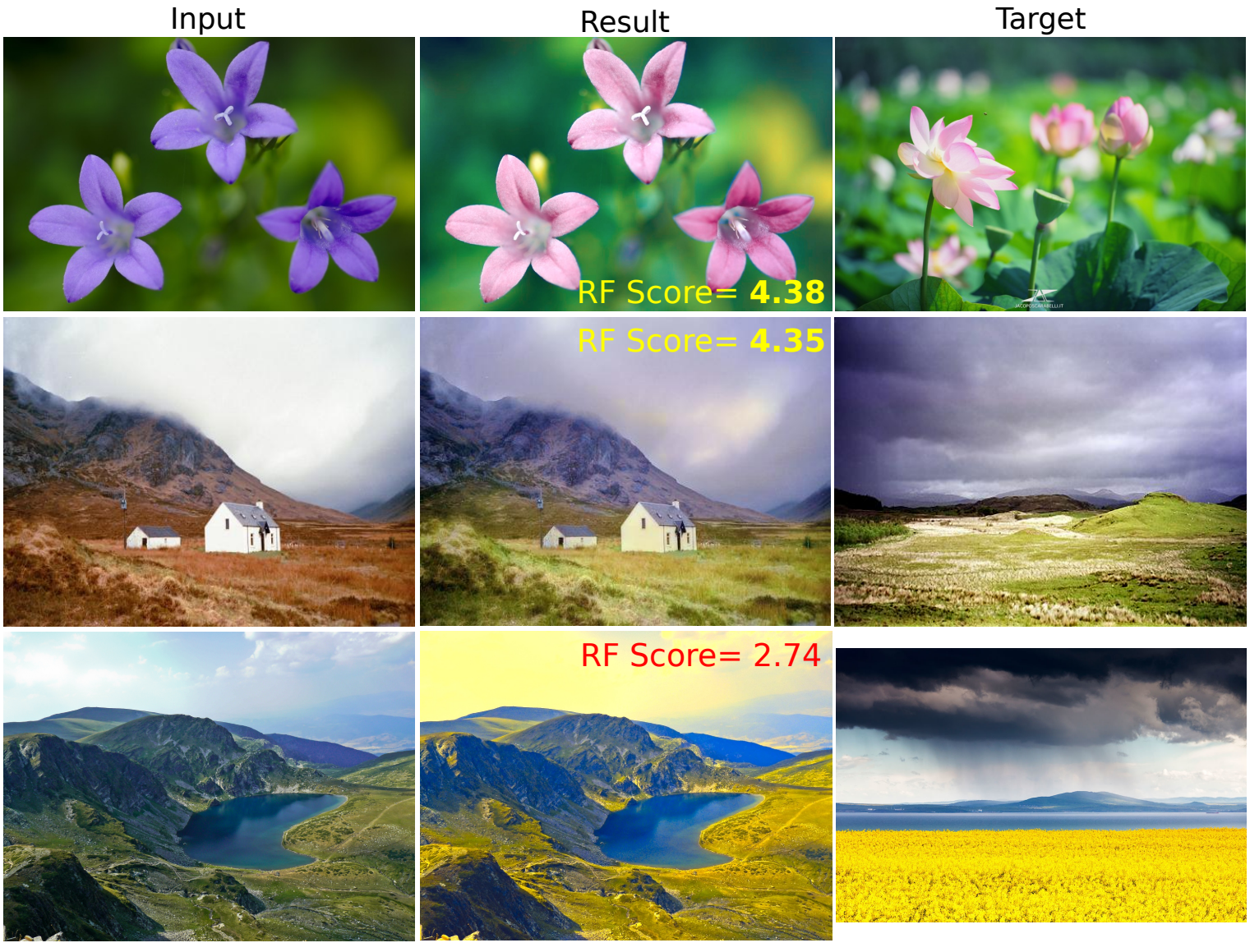
Flowchart


Supervised learning
Prediction
(input, target, result)
Input image: I
Target image: J
Result: R
Subjective user study
Feature extraction
Feature extraction
Set of subjective scores
Regression
Random forests
Model
Prediction score:
3.70

Results



Results



Conclusion

Advantages

- The first attempt to compute a perceptual metric considering an ensemble of objective metrics;
- Designed to evaluate the quality of color transfer methods.

Conclusion

Advantages

- The first attempt to compute a perceptual metric considering an ensemble of objective metrics;
- Designed to evaluate the quality of color transfer methods.

Limitations

- Not suitable for evaluating style transfer methods;
- Not suitable for methods which modify texture or use paintings/sketches as target images.

References I

Bonneel N., Sunkavalli K., Paris S., Pfister H.:
Example-based video color grading.
ACM Trans. Graph. 32, 4 (2013), 39–1.

Fairchild M. D., Johnson G. M.:
icam framework for image appearance, differences, and quality.

Journal of Electronic Imaging 13, 1 (2004), 126–138.

Faridul H. S., Pouli T., Chamaret C., Stauder J., Trémeau A., Reinhard E., et al.:
A survey of color mapping and its applications.
In *Eurographics (State of the Art Reports)* (2014), pp. 43–67.

References II

Frigo O., Sabater N., Demoulin V., Hellier P.:
Optimal transportation for example-guided color transfer.
In *Asian Conference on Computer Vision* (2014), Springer, pp. 655–670.

Hristova H., Le Meur O., Cozot R., Bouatouch K.:

Style-aware robust color transfer.
In *Proceedings of the workshop on Computational Aesthetics* (2015), Eurographics Association, pp. 67–77.

Le Meur O., Baccino T.:
Methods for comparing scanpaths and saliency maps: strengths and weaknesses.
Behavior research methods 45, 1 (2013), 251–266.

References III

Shih Y., Paris S., Durand F., Freeman W. T.:
Data-driven hallucination of different times of day from a single outdoor photo.
ACM Transactions on Graphics (TOG) 32, 6 (2013), 200.

Questions?

Thank you for your attention!

For more information, please visit

http://people.irisa.fr/Hristina.Hristova

Accuracy

Coefficient of determination

- Coefficient of determination: 0.94;

Accuracy

Coefficient of determination

- Coefficient of determination: 0.94;
- A large portion of the variability of the subjective scores (user study) - explained by our model.

Accuracy

Coefficient of determination

- Coefficient of determination: 0.94;
- A large portion of the variability of the subjective scores (user study) - explained by our model.

Out-of-bag (OOB) error

- OOB sample set - comprised of all the bootstrap datasets which do not contain a particular feature;

Accuracy

Coefficient of determination

- Coefficient of determination: 0.94;
- A large portion of the variability of the subjective scores (user study) - explained by our model.

Out-of-bag (OOB) error

- OOB sample set - comprised of all the bootstrap datasets which do not contain a particular feature;
- OOB error - the generalization error over the OOB samples;

Accuracy

Coefficient of determination

- Coefficient of determination: 0.94;
- A large portion of the variability of the subjective scores (user study) - explained by our model.

Out-of-bag (OOB) error

- OOB sample set - comprised of all the bootstrap datasets which do not contain a particular feature;
- OOB error - the generalization error over the OOB samples;
- OOB error: 0.58.