

Psychovisual Perception Scale Based on a Neural Network

Vladimir Budak ^[0000-0003-4750-0160] and Ekaterina Ilyina ^[0000-0003-0783-0931]

Moscow Power Engineering Institute (National Research University),
Krasnokazarmennaya 14, Moscow, Russia
budakvp@gmail.com, kitesika@gmail.com

Abstract. The purpose of this article is to construct a psychophysical scale of visual perception from lighting scene based on a direct propagation neural network using for assessment of real or synthesized images with spatial brightness distribution.

Visual perception assessments of different scenes were obtained for 10 observers at the experimental installation of the Department of lighting engineering of the MPEI (NRU). These results were checked and found out agreed with the numerical scale of visual perception proposed by Lekish and Holladay. Neural network was trained to predict a sensation at the level of 40-70%, depending on the scale category. For more careful prediction level in each of 5 categories of scale a new experiment should be done with new calibration and with tested instructions and with more observers involved.

The novelty consists in using a neural network as an expert to assess the degree of comfort of the lighting scene.

Keywords: Glare Discomfort, Scale of Visual Perception, Neural Network

1 Introduction

Currently, neural networks (NN) are widely used in image and text recognition tasks, in medicine as diagnostic systems, in quality control systems and in many other fields where it is impossible to consider all the conditions that affect the decision. And therefore, only the most important ones are considered. In lighting engineering, the problem of psychophysical assessment of a visual perception of the spatial brightness distribution into person's field of view can be assigned to this class of tasks. Processing of visual information can be conditionally represented as: "Stimulus – sensation – perception - reaction" where the stimulus is the spatial brightness distribution and the reaction is the sensation of comfort or discomfort caused by stimulus.

Modern lighting calculation programs allow us to get quite realistic visualizations of the designed lighting installations. To evaluate the quality of the brightness distribution

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

in terms of visual comfort using real or synthesized images, a psychophysical lighting comfort scale (PLCS) should be built to rank all these images. PLCS could be used to set the relationship between human sensation and the numerical values of the light source's and background's brightness on the image. The probabilistic model of transition from category to category of sensation should be used in PLCS to predict with some degree of accuracy what reaction can be expected from a lighting scene.

The purpose of this work is to construct a PLCS using empirical data obtained from an experiment where observer's reaction the depending on the brightness of a bright light source observed on a uniform background is evaluated when performing a typical visual task.

Lekish and Holladay were the first who used a psychophysical assessment of visual perception of lighting [1] and tried to find the relationship between the numerical values of source's and background's brightness, and a set of human responses, expressed in the categories of lighting comfort. They suggested a scale like as is scarcely noticeable, is most pleasant, is still pleasant, is at limit of pleasant, is very comfortable, is still comfortable, is less comfortable, is at the boundary between comfort and discomfort, is perceptibly uncomfortable, is thoroughly uncomfortable, is at boundary between objectionable and intolerable, is irritating and is painful. They had got an empirical formula for value K as measure of «sensation» that looks like this:

$$K = \log(L_s) + 0.25\log(Q) - 0.3\log(L_{ad}) \quad (1)$$

where L_s –brightness of source, Q – solid angle of light source, L_{ad} – brightness of adaptation or background for source. Each numeric value of K corresponds to one of the thirteen categories of the scale, for example, $K = 1.9$ indicates a feeling on the boundary between comfort and discomfort (BCD).

In turn, Hopkinson's research [2] showed that sensations on the scale could not be clearly defined as values of brightness, they can only be expressed by an interval. So the transition from scale's category to another can be set by a probability function. For example, Hopkinson shows that the function describing BCD has form as a sigmoid. Thus, judge a sensation can be predicted only with a certain probability.

The scale of sensations was compared with the values of the discomfort index calculated using an empirical formula for a single source of light using a system of categorical judgments [3]. Since the sensitivity of the eye changes exponentially as the brightness level increases, the standard deviation interval also increases for each subsequent interval on the scale. Based on this, the hypothesis of a normal distribution of ratings on the scale was confirmed, which allows us to find the numerical boundaries of the intervals of subjective categories in standard deviation units and determine the numerical boundaries of the intervals for a certain probability value, for example, $p=0.5$.

Construction of a PLCS using the probabilistic model of transition from category to category can give a more complete description how the light environment (the brightness distribution in space) affects a person. It can become a measure of the typical attitude of a group of observers to lighting conditions.

Lakish's and Guth's experiment (in 1949) investigated the influence of various factors on the visual perception on BCD for different brightness levels of glare source in

the range from 1000 to 6000 cd/m², and as a result, these studies contributed to clarifying definition, which formed the basis of modern unfired glare ratings (UGR) [4].

An experimental at the Department of lighting engineering in MPEI (NRU) was built to study the discomfort from glare sources with various shapes and brightness in the observer's field of view/ It's similar to the Lakish's and Guth's installation, but has modern light sources and a wider range of brightness from 33 to 100,000 cd/m². Modern experiment for BDC shown that the results were agreed with the experiment of Lekish and Guth installation [5]. Initially, the psychophysical scale proposed by Lekish and Holladay was used to find BCD. But later a scale was reduced from 13 to 5 categories: barely noticeable, comfortable, uncomfortable, unpleasant, and painful. From our point of view these 5 categories are enough for building of new PLCS.

From the point of view of NN, the task of PLCS's construction is most like the class of task «many-to-many», which include image recognition. Such a tasks have several input parameters, and the result of the NN is the classification of objects into several categories. When a network divides input vectors into two classes, it is sufficient to have only one network output that takes the values «1» or «0». But in our case, the scale of sensations has 5 categories, so the number of outputs should be equal to 5. For example if the observer's response corresponds to a category barely noticeable then the value should be written as "1", the rest as "0", and this logic of recording should be applied to each observer's response.

As a result, the problem is reduced next conditions: there are several photos of the lighting system with a spatial distribution of brightness at the input, and there are real estimates of the observers of these scenes corresponding to the five categories of the scale at the output. During training, the NN will extract certain features from the input images, collect and classify them in a certain way. After training, the network should output the probability of an image or group of images falling into each category of the scale. The maximum probability will determine the category of the comfort scale, that is, the most likely reaction of typical observers to the distribution of brightness in the room.

2 Methods

If we simplify the distribution of brightness on image to the ratio of the brightness of light source of finite size viewed on the uniform background, then the input vector into NN have two variables: $y_1 = L_S$ and $y_2 = L_{ad}$.

In the future, the length of the vector y can be changed by adding, for example, correlated color temperature of the light source (y_3). Moreover, later y_1 and y_2 can be used both from experiment and from an image with a spatial distribution of brightness in a real photograph or modeled in a lighting calculated program. The recognizing and classification of images tasks are well solved using convolutional NN. In the framework of this article, we will not consider this issue.

The input and output data for training of the NN were obtained from an experiment performed on an experimental installation at the Department of Lighting Engineering of the MPEI (NRU). The technique consisted of assessing the sensation when a glare

source appeared in the observer's field of vision while reading the inscription. The background brightness was uniform. The adaptation brightness was taken equal to the background brightness. The experiment involved 7 observers with a background brightness of 57 and 113 cd/m^2 in the dark time, and 3 observers with a background brightness of 66, 77, 95 and 119 cd/m^2 in the daytime in cloudy weather. Each observer evaluated the first sensation and registered the value of source brightness. Then the brightness of the source changed (gradually increased or decreased) till the observer registered the second sensation and these actions were repeated up to fifth sensation. Then the background brightness was changed to next. The measurement was carried out both on an ascending and descending scale. The observer was seated on the chair at a distance of 1.5 meters from the inscription "Experiment", which he read. The inscription was at the level of the line of sight. The source with diameter 30 mm was located just above the line of sight.

Reading of inscription imitates condition for performing visual work when glare source can appear in the person's field of view and cause sensation of discomfort. Since the scale of sensations refers to the scale of verbal categories, the observer can find it difficult to spread his feelings in a wide range. That's why special attention was paid to the briefing. It should be clear to the observer in terms that are familiar to him. In our experiment the briefing was as follows: barely noticeable - the source becomes noticeable during reading the inscription, this at the moment when the observer detects the source at the first time; "comfortable" - it is pleasant to read the inscription and the light source does but does not interfere with reading, "uncomfortable" - the sensation of a "buzzing fly", the source already attracts attention, but does not, "unpleasant" - distracts from reading the inscription, "Unbearable" means painfully.

Initially, the experiment was done with LEDs covered white fabric without limiting the angular size of the source. Later it was fixed. It is very important to limit angular size of the source during experiment because it can affect the results.

The installation was calibrated for each the brightness of source and background using the Konica Minolta LS-110 brightness meter. A feature of the installation is that the brightness measurement is carried out indirectly. The observer sets not the source's brightness but the voltage on the LEDs (U). The calibration curve $LS = f(U)$ shown that in the range from 7.2 to 7.8 V, the brightness increases from 20 to 1870 cd/m^2 (see Fig.1).

Brightness of background is set by the level of general lighting in the room. When it turned out there is a contribution from general lighting into the brightness of source surface. Therefore, calibration was performed for each background of brightness level separately, in order to reduce the error at low brightness levels. This limitation affects the ability to accurately determine the interval of brightness for sensations: barely noticeable and comfortable.

The measurement was carried out at one adaptation level, then it was changed. There was a pause between the measurements so that the observer's eyes could adapt. All measured and registered data were converted into input and output vectors for transferring them into nprtool in Matlab. This application helps to solve the problems of classification of pattern recognition [6]. After that, a NN with direct distribution was created. It contained several hidden layers and an output layer (see Fig. 2). All neurons

may have their own activation function; the number of neurons corresponds to the number of classes n on the last layer, in our case $n = 5$.

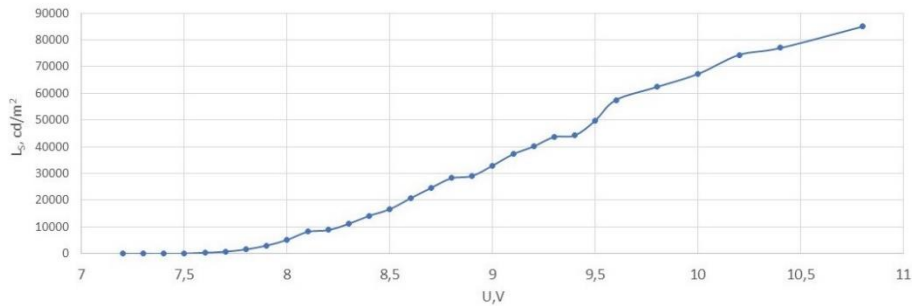


Fig. 1. Calibration curve for source brightness from the voltage on the LED at $L_{ad} = 54 \text{ cd/m}^2$

The training of NN was performed using the standard *trainscg* function, which can train any NN if its weight, clean input signal and transfer functions have derivative functions. To solve multidimensional logical problems, the Softmax activation function is applied, which looks like this:

$$S(y)_i = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}} \quad (2)$$

The function converts a vector with dimension j into a vector S with the same dimension. Each coordinate $S(y)_i$ of the resulting vector is represented by a real number in the interval $[0,1]$ and the sum of coordinates $S(y)_i$ is equal to 1. The layer index n is usually omitted, therefore which implies that this is the last layer.

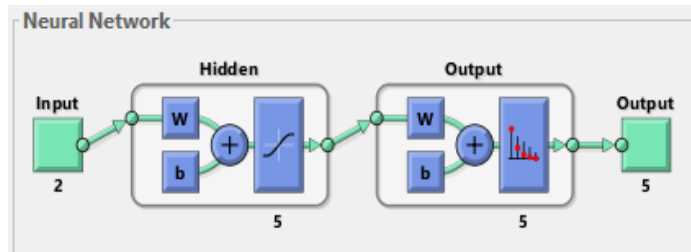


Fig. 2. The scheme of the neural network. (Input - input vector of dimension $j = 2$, Hidden - number of hidden layers $n = 5$, Output - output layer with 5 classes, Output - probabilities for each class)

Our NN has five outputs, then it uses Softmax function to produce a set of probabilities for each sensation caused by the brightness ratio. To assess the reliability, we use the surface defined as:

$$p = f(L_s, L_{ad}) \tag{3}$$

where p is the maximum probability of each event on the PLCS. In other words, this probability surface shows how each point in space with coordinates L_s and L_{ad} refers to a sensation. The higher the probability value, the higher the confidence. To determine the sensations on the PLCS, we build a sensation surface in the same space, which can be defined as:

$$R = f(L_s, L_{ad}) \tag{4}$$

Thus, our model of scale may represent some surface - a field of predictions for a given initial data.

3 Results

Average brightness values obtained during the experiment for each category of the scale were used to calculate values K using formula (1). Calculated values K were matched with Lekish-Holliday's scale (see Fig. 3).

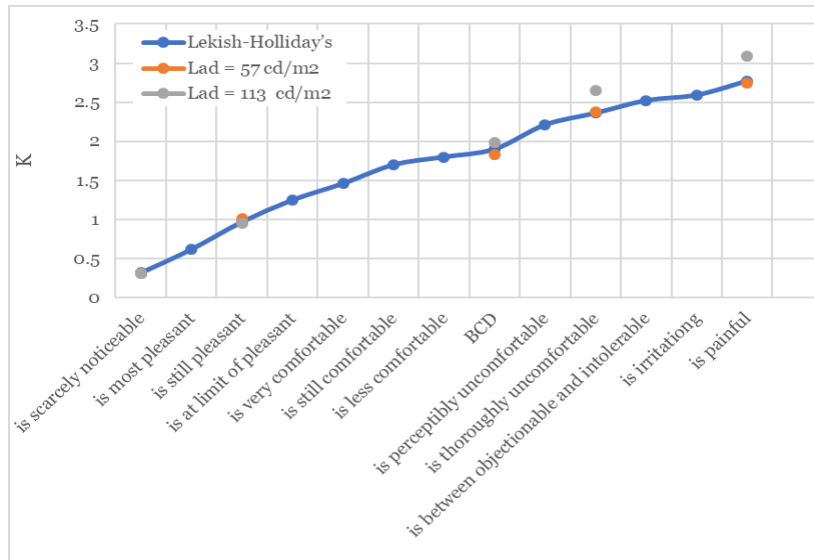


Fig. 3. Value K and the corresponding sensation on the Lekish-Halladay's scale

It can be noted that the values K for *barely noticeable* and *still nice coincide* rather good. There is a slight discrepancy for the category of *BCD. Thoroughly uncomfortable*

and *painful coincide* at $L_{sd} = 54 \text{ cd/m}^2$ and slightly diverge at $L_{sd} = 113 \text{ cd/m}^2$. In this case, thoroughly uncomfortable corresponds to the instruction unpleasant in our experiment. The all values $K > 2.7$ are considered as *painful* [1]. Nevertheless, such a graph does not provide information how the feeling can be translated into the brightness of source and background and how to determine the width of each interval. To approximate the sensation according to the normal law using response frequencies for a five-point scale, it's needed to find the only numerical value for one fixed brightness of the source and background. We constructed graphs with the distribution of response frequencies for each category of the scale (see Fig. 4). Several maxima can be observed within each category. This indicates the lack of statistical data. In *barely noticeable* and *comfortable* the maxima are shifted toward the light source's brightness 42 and 450 cd/m^2 respectively. For *uncomfortable* there are two maxima 3107 and 8910 cd/m^2 instead of one. *Unpleasant* and *unbearably* have a more smeared character without a clear maximum. Presumably it could be due to insufficient data or because it was difficult to correctly separate the person's sensation of *discomfort* and *unpleasant*. As for the *painful* category, everyone has their own threshold and more statistics are needed here.

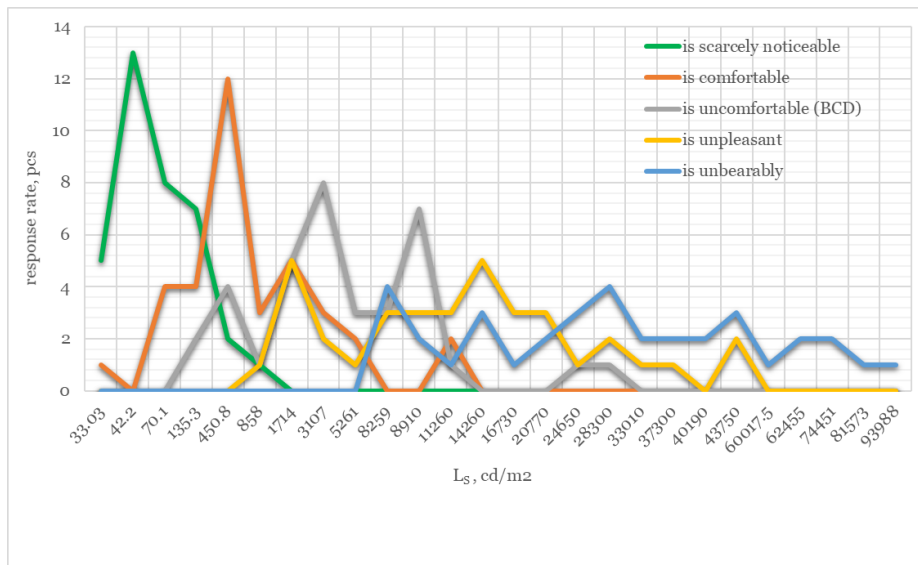


Fig. 4. Graphs of the distribution of response frequencies for each category of the scale depending on the brightness of the light source

All the results were used to train NN and to construct the PLCS where sensation is a function of brightness of the source and background. Having built a multilayer artificial NN it's important to determine whether it adequately describes the data and the dependencies between them. As a rule the number of neurons in the hidden layer is selected experimentally by comparing the accuracy of NN. So, the number of hidden layers in the NN is a custom parameter. The minimum values of cross entropy (CE) and

percent of errors (the proportion of incorrectly classified examples) are preferred for the training, test, and test sets. The smaller the error, the better result[10].

In our test, we take the number of hidden layers from 2 to 100. A graph of the error depending on the hidden layers of the NN, has a form like a cardiogram. Anyway it was determined that the NN with 32 layers has minimum percentage of error, and NN with 11 layers has an error is closest to the minimum.

Not only the CE parameters and the percent of errors could be used to evaluate the quality of the NS operation, but also a Confusion Matrix is used. For each class of observations, the results of assigning observations to a class are given. The matrix allows to see whether the classifier confuses classes. The matrix columns correspond to the predicted classes, and the rows correspond to the actual classes.

Fig.5 shows the percentage of error for NN with 32 layers. Classes from 1 to 5 mean categories on the scale from *barely noticeable* to *unbearable*. According to the Confusion Matrix, grades from 1 and 5 have a minimum number of errors 9.8% and 28.6% respectively. These values correspond to the criterion of prediction accuracy of at least 70%. Grades 2, 3 and 4, have a lower the prediction accuracy. Let's see how these parameters change for NN with 11 layers (see Fig.5). The percentage of errors in all classes except the grades 2 and 4 decreases. In this case, accuracy for grade 2 (*comfortable*) and grade 4 (*unpleasant*) is rather low. But for grade 3 (*discomfort*) it is over 60% which is quite high. This clearly indicates that the NN with 11 layers works more adequately compared with NN with 11 layers. To improve the accuracy of the prediction, additional data from the experiment is required for at least the *comfortable* and *unpleasant* categories.

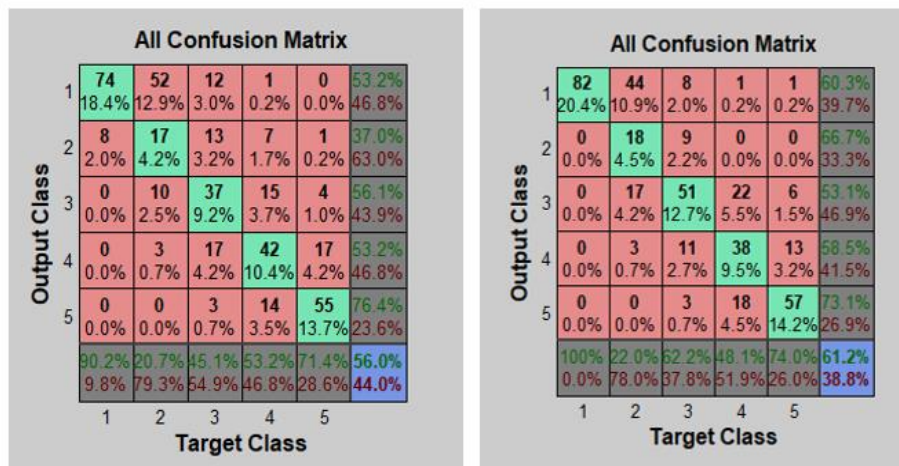


Fig. 5. All Confusion Matrix for 32 and 11 layers

The shape of the Receiver Operating Characteristic (ROC-curves) is also an important indicator of the quality of the neural network. For an ideal classifier, the ROC-curve graph passes through the upper-left corner, where the percentage of true positive cases is 100% or 1, and the percentage of false positive examples is zero. Therefore,

the closer the curve is to the upper-left corner, the higher the predictive power of the model. On the contrary, the smaller the curve bends and the closer it is to the diagonal line, the less efficient.

As it can be seen at Fig.6 neural network with 11 layers has the ROC-curves that pass closer to the upper-left corner, which means that the predictive ability of this model is higher than for other two. At the same time, it is clearly noticeable that ROC-curves for model with 32 layers passes below the diagonal. It means that model lie.

Let's construct surfaces p and R defined by formulas (3) and (4) for L_S range from 100 to 100000 cd/m^2 and L_{ad} range from 54 to 120 cd/m^2 using NN with 38 and 11 hidden layers. From Fig.7 the shapes of both surfaces change depending on the number of layers. The surface obtained based on NN with 11 layers looks less ragged.

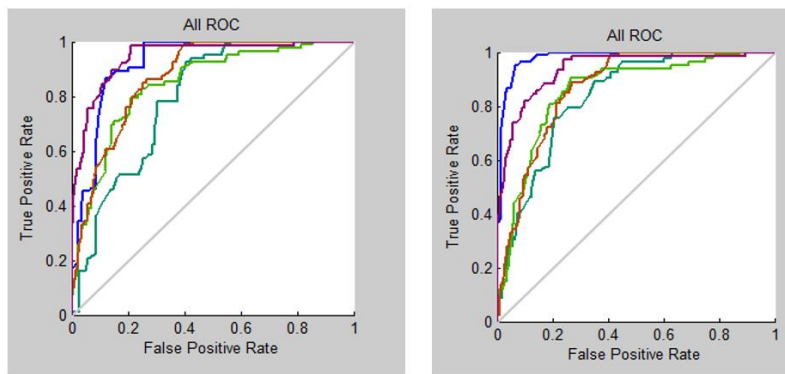


Fig. 6. ROC-curves depending on the number of hidden layers: 32(left) and 11 (right)

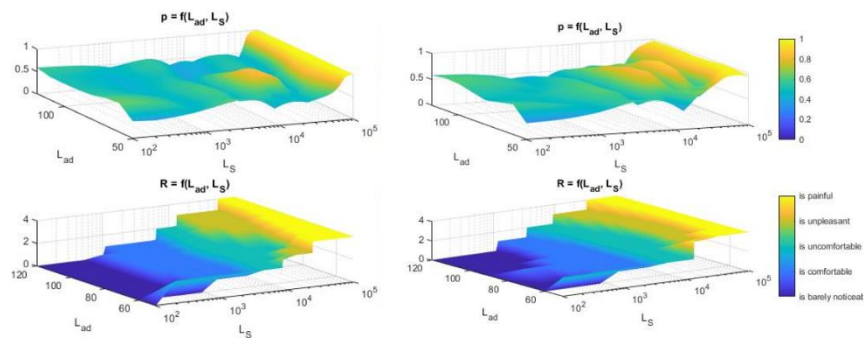


Fig. 7. Probability surfaces for PLCS for cases: NN with 38 layers (left) and NN with 11 layers (right) respectively

The higher the accuracy of the prediction (or probability), the more NN works like an "expert" who can predict: what kind of sensation will be caused by the ratio of the brightness of source and the background in a typical group of observers.

4 Discussions

In this article, we reviewed the results of an experiment on MPEI's installation and the idea of using neural networks to construct PLCS depending on the ratio of the brightness of source and the background was tested.

Despite that response rates (frequencies) curves (Fig. 4) have no maxima for unpleasant and unbearable, the probability surface p shows the greatest NN's accuracy for these categories. This may be because the observer can easily detect them, that's why NN can be training more accurately and provide classification with $p > 0.7$. As for barely noticeable, the surface p shows the probability $p > 0.5$ (green and yellow zone). Comfortable and uncomfortable have the lowest values of p , they contain zones where $p < 0.5$ (blue color). Thus, the resulting model has a prediction accuracy of 40-70%, depending on the category of scale.

This work has shown that this idea can be used to build a model for psychological assessment of the spatial distribution of brightness in a lighting scene from comfort point of view if provided a enough experimental sample for training. Even though now PLCS has lower accuracy than required, this work allows us to formulate criteria for the future model of the scale and requirements for a new experiment. To provide best predicational ability of NN the brightness of background in new experiment should be at the range from 50 to 200 cd/m² with step of 10 cd/m² and a greater number of observers should be involved.

Also, instructions used in this work for calibration of the experimental installation and for briefing of observers before starting can be used in future.

This can improve the accuracy of the input data, and therefore improve the work of the NN as the main "expert" that assess the comfort of lighting scene using synthesized images.

References

1. Luckiesh, M., Holladay, L.: Glare and Visibility. Transactions of the IES. 20: 221(1925)
2. Hopkinson, R.G.: Evaluation of Glare. Illuminating Engineering 52(6), 305–321 (1957)
3. Kotik, G., Matveev? A., Perejma, V., Tohadze, I.: Kategorijnye ocenki kachestva i ih vzaimosvyaz' na psihofizicheskoy shkale, Svetotekhnika, 3, 3-5 (1975)
4. Luckiesh, M., Guth, S.: Brightness in the visual field at borderline between comfort and discomfort. Illuminating Engineering, 44(11), 650 (1949)
5. Budak, V., ZHeltov, V., Meshkova, T., Notfullin, R.: Ocenka kachestva osveshcheniya na osnove prostranstvenno-uglovogo raspredeleniya yarkosti. Svetotekhnika, 3, 17-22 (2017)
6. Beale, M., Hagan, M., Demuth, H.: Deep Learning Toolbox Getting Started Guide R2019b, MathWorks, 2019, 162