A. Implementation Details

For the single-view 3D reconstruction experiment, we closely orient ourselves on the setup by Liu *et al*. [1]. We use the same model architecture [1] and also train with a batch size of 64 for 250 000 steps using the Adam optimizer [2] We also schedule the learning rate to 10^{-4} for the first 150 000 steps and use a learning rate of $3 \cdot 10^{-5}$ for the remaining training. At this point (after the first 150 000 steps), we also decrease the temperature τ by a factor of 0.3.

Using different learning rates (as an ablation) did not improve the results.

B. Distributions

In this section, we define each of the presented distributions / sigmoid functions. Figure 5 displays the respective CDFs and PDFs.

Note that, for each distribution, the PDFs f is defined as the derivative of the CDF F. Also, note that a reversed (Rev.) CDF is defined as $F_{\text{Rev.}}(x) = 1 - F(-x)$, which means that $F_{\text{Rev.}} = F$ for symmetric distributions. The square-root distribution F_{sq} is defined in terms of F as in Equation (5). Therefore, in the following, we will define the distributions via their CDFs F.

Heaviside

$$x \mapsto \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

Uniform

$$x \mapsto \begin{cases} 0 & \text{if } x < -1 \\ 0.5 \cdot (1 + x) & \text{if } -1 \leq x \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

Cubic Hermite

$$x \mapsto \begin{cases} 0 & \text{if } x < -1 \\ 3y^2 - 2y^3 & \text{if } -1 \leq x \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{8}$$

where $y := (x + 1)/2$.

Wigner Semicircle

$$x \mapsto \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{2} + \frac{x\sqrt{1-x^2}}{\pi} + \frac{\arcsin(x)}{\pi} & \text{if } -1 \leq x \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

Gaussian

$$x \mapsto \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{10}$$

Laplace

$$x \mapsto \begin{cases} \frac{1}{2}\exp(x) & \text{if } x \leq 0 \\ 1 - \frac{1}{2}\exp(-x) & \text{if } x \geq 0 \end{cases} \tag{11}$$

Logistic

$$x \mapsto \frac{1}{1 + \exp(-x)} \tag{12}$$

Hyperbolic secant / Gudermannian

$$x \mapsto \frac{2}{\pi}\arctan\left(\exp\left(\frac{\pi}{2}x\right)\right) \tag{13}$$

Cauchy

$$x \mapsto \frac{1}{\pi}\arctan(x) + \frac{1}{2} \tag{14}$$

Reciprocal

$$x \mapsto x/(2 + 2|x|) + 1/2 \tag{15}$$

Gumbel-Max

$$x \mapsto e^{-e^{-x}} \tag{16}$$

Gumbel-Min

$$x \mapsto e^{-e^{x}} \tag{17}$$

Exponential

$$x \mapsto 1 - e^{-x} \tag{18}$$

Levy

$$x \mapsto 2 - 2\Phi\left(\sqrt{\frac{1}{x}}\right) \tag{19}$$

where Φ is the CDF of the standard normal distribution.

Gamma

$$x \mapsto \frac{1}{\Gamma(p)}\gamma(p, x) \tag{20}$$

where $\gamma(p, x)$ is the lower incomplete gamma function and $p > 0$ is the shape parameter.

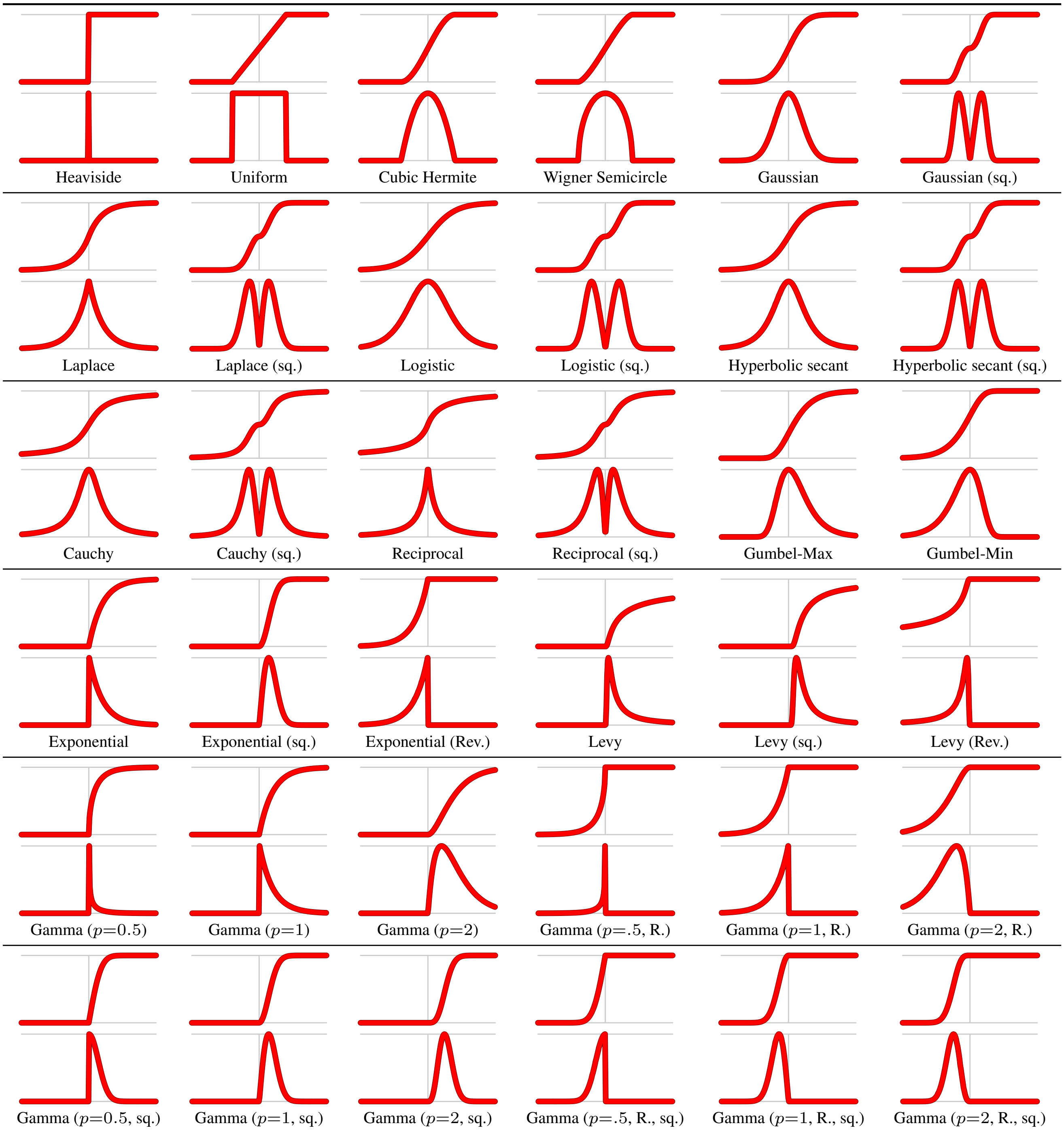

Table 5. Visualization of CDFs (top) and PDFs (bottom) for different distributions.

C. T-Norms and T-Conorms

The axiomatic approach to multi-valued logics (which we need to combine the occlusions by different faces in a "soft" manner) is based on defining reasonable properties for truth functions. We stated the axioms for multi-valued generalizations of the disjunction (logical "or"), called T-conorms, in Definition 2. Here we complement this with the axioms for multi-valued generalizations of the conjunction (logical "and"), which are called T-norms.

Definition 6 (T-norm)**.** *A T-norm (triangular norm) is a binary operation* $\top : [0,1] \times [0,1] \to [0,1]$*, which satisfies*

- *associativity:* $\top(a, \top(b,c)) = \top(\top(a,b),c)$*,*
- *commutativity:* $\top(a,b) = \top(b,a)$*,*
- *monotonicity:* $(a \le c) \wedge (b \le d) \Rightarrow \top(a,b) \le \top(c,d)$*,*
- 1 *is a neutral element:* $\top(a,1) = a$*.*

Clearly these axioms ensure that the corners of the unit square, that is, the value pairs considered in classical logic, are processed as with a standard conjunction: neutral element and commutativity imply that $(1,1) \mapsto 1$, $(0,1) \mapsto 0$, $(1,0) \mapsto 0$. From one of the latter two and monotonicity it follows $(0,0) \mapsto 0$. Analogously, the axioms of T-conorms ensure that the corners of the unit square are processed as with a standard disjunction. Actually, the axioms already fix the values not only at the corners, but on the boundaries of the unit square. Only inside the unit square (that is, for $(0,1)^2$) T-norms (as well as T-conorms) can differ.

Minimum	$\top^M(a,b) = \min(a,b)$
Probabilistic	$\top^P(a,b) = ab$
Einstein	$\top^E(a,b) = \frac{ab}{2-a-b+ab}$
Hamacher	$\top^H_p(a,b) = \frac{ab}{p+(1-p)(a+b-ab)}$
Frank	$\top^F_p(a,b) = \log_p\left(1 + \frac{(p^a-1)(p^b-1)}{p-1}\right)$
Yager	$\top^Y_p(a,b) = \max\left(0, 1-((1-a)^p + (1-b)^p)^{\frac{1}{p}}\right)$
Aczél-Alsina	$\top^A_p(a,b) = \exp\left(-(\lvert\log(a)\rvert^p + \lvert\log(b)\rvert^p)^{\frac{1}{p}}\right)$
Dombi	$\top^D_p(a,b) = \left(1 + \left(\left(\frac{1-a}{a}\right)^p + \left(\frac{1-b}{b}\right)^p\right)^{\frac{1}{p}}\right)^{-1}$
Schweizer-Sklar	$\top^S_p(a,b) = (a^p + b^p - 1)^{\frac{1}{p}}$

Table 6. (Families of) T-norms.

In the theory of multi-valued logics, and especially in fuzzy logic [3], it was established that the largest possible T-norm is the minimum and the smallest possible T-conorm is the maximum: for any T-norm $\top$ it is $\top(a,b) \le \min(a,b)$ and for any T-conorm $\perp$ it is $\perp(a,b) \ge \max(a,b)$. The other extremes, that is, the smallest possible T-norm and the largest possible T-conorm are the so-called drastic T-norm, defined as $\top^\circ(a,b) = 0$ for $(a,b) \in (0,1)^2$, and the drastic T-conorm, defined as $\perp^\circ(a,b) = 1$ for $(a,b) \in (0,1)^2$. Hence it is $\top(a,b) \ge \top^\circ(a,b)$ for any T-norm $\top$ and $\perp(a,b) \le \perp^\circ(a,b)$ for any T-conorm $\perp$. We do not consider the drastic T-conorm for an occlusion test, because it clearly does not yield useful gradients.

As already mentioned in the paper, it is common to combine a T-norm $\top$, a T-conorm $\perp$ and a negation N (or complement, most commonly $N(a) = 1 - a$) so that DeMorgan's laws hold. Such a triplet is often called a *dual triplet*. In Tables 6 and 7 we show the formulas for the families of T-norms and T-conorms, respectively, where matching lines together with the standard negation $N(a) = 1 - a$ form dual triplets. Note that, for some families, we limited the range of values for the parameter p (see Table 2) compared to more general definitions [3].

C.1. T-conorm Plots

Figures 7 and 8 display the considered set of T-conorms.

Maximum	$\perp^{M}(a,b) = \max(a,b)$
Probabilistic	$\perp^{P}(a,b) = a + b - ab$
Einstein	$\perp^{E}(a,b) = \perp_2^{H}(a,b) = \frac{a+b}{1+ab}$
Hamacher	$\perp_p^{H}(a,b) = \frac{a+b+(p-2)ab}{1+(p-1)ab}$
Frank	$\perp_p^{F}(a,b) = 1 - \log_p\left(1 + \frac{(p^{1-a}-1)(p^{1-b}-1)}{p-1}\right)$
Yager	$\perp_p^{Y}(a,b) = \min\left(1, (a^p + b^p)^{\frac{1}{p}}\right)$
Aczél-Alsina	$\perp_p^{A}(a,b) = 1 - \exp\left(-(\lvert\log(1-a)\rvert^p + \lvert\log(1-b)\rvert^p)^{\frac{1}{p}}\right)$
Dombi	$\perp_p^{D}(a,b) = \left(1 + \left(\left(\frac{1-a}{a}\right)^p + \left(\frac{1-b}{b}\right)^p\right)^{-\frac{1}{p}}\right)^{-1}$
Schweizer-Sklar	$\perp_p^{S}(a,b) = 1 - ((1-a)^p + (1-b)^p - 1)^{\frac{1}{p}}$

Table 7. (Families of) T-conorms.

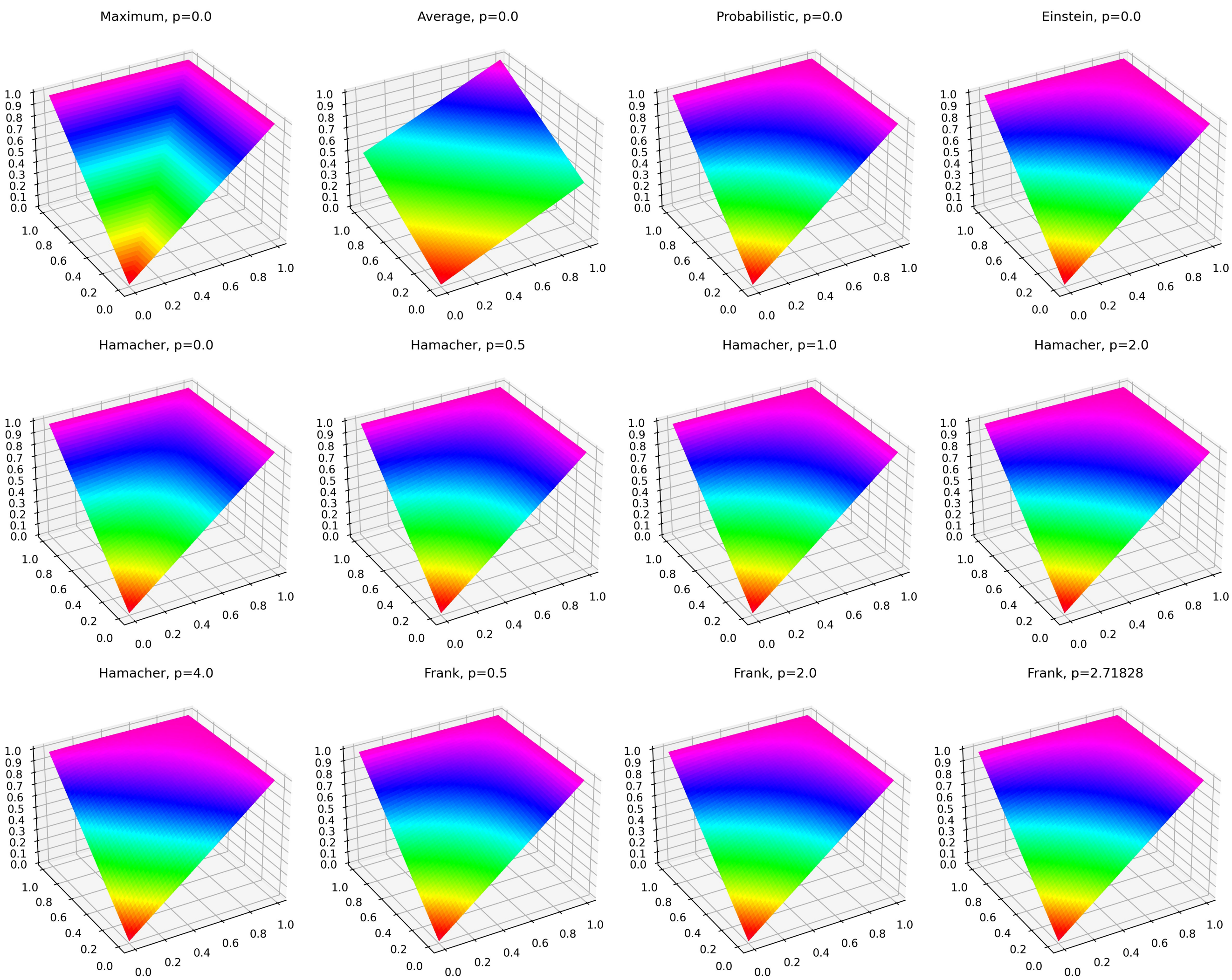

Figure 7. T-conorm plots (1/2). Note that 'Average' is not a T-cornom and just included for reference. Also, Note how 'Probabilistic' is equal to 'Hamacher $p = 1$' and 'Einstein' is equal to 'Hamacher $p = 2$'.

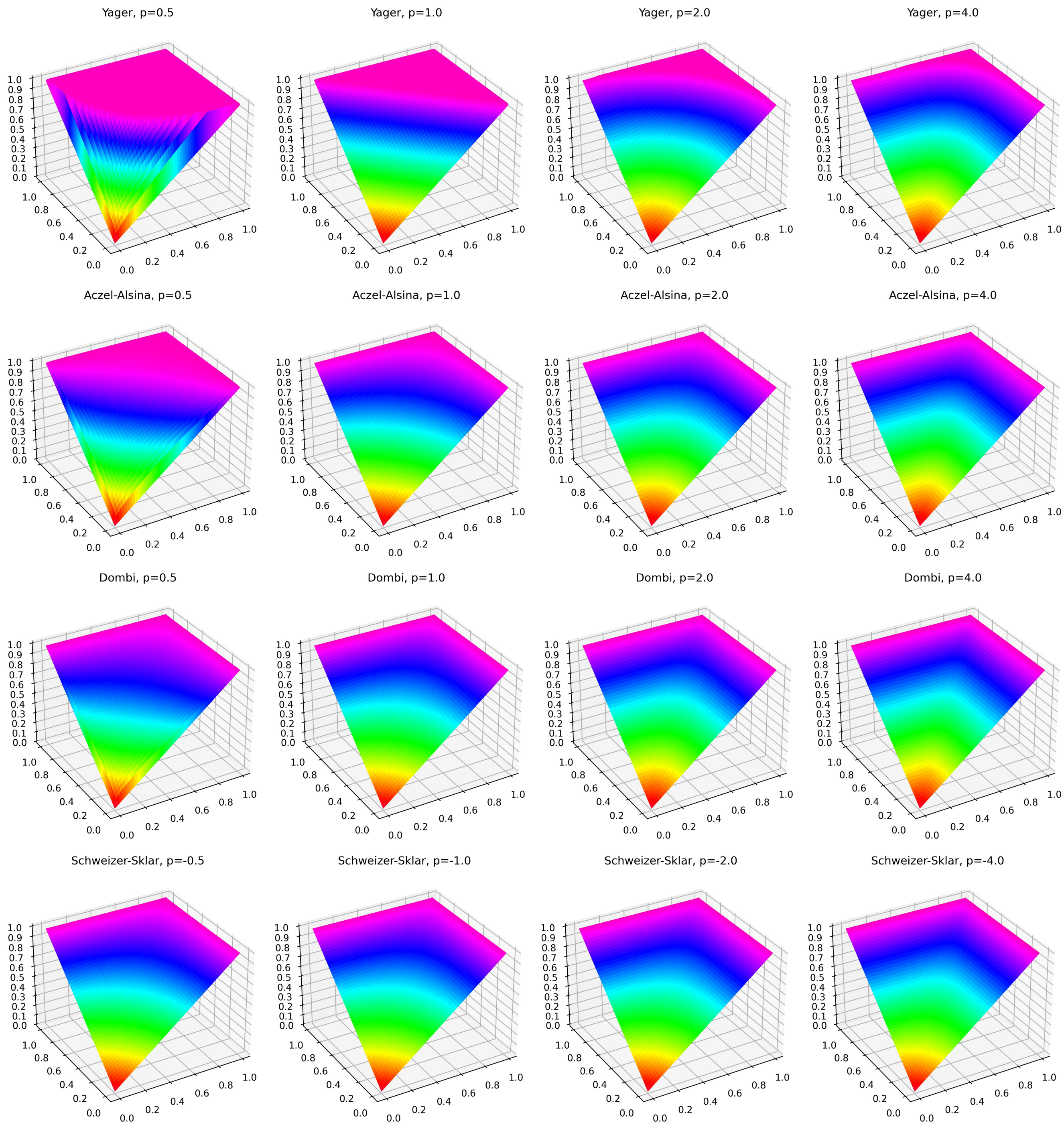

Figure 8. T-conorm plots (2/2).

D. Additional Plots

See Figures 9 and 10.

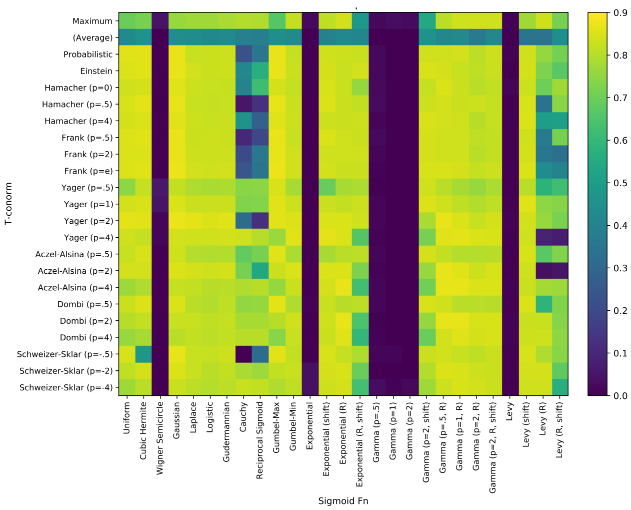

Figure 9. Results for the tea pot camera pose optimization task for the respective square-root distribution F_{sq}.

References

[1] S. Liu, T. Li, W. Chen, and H. Li, "Soft Rasterizer: A Differentiable Renderer for Image-based 3D Reasoning," in *Proc. International Conference on Computer Vision (ICCV)*, 2019.

[2] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[3] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic — Theory and Applications*. Upper Saddle River, New Jersey: Prentice Hall, 1995.

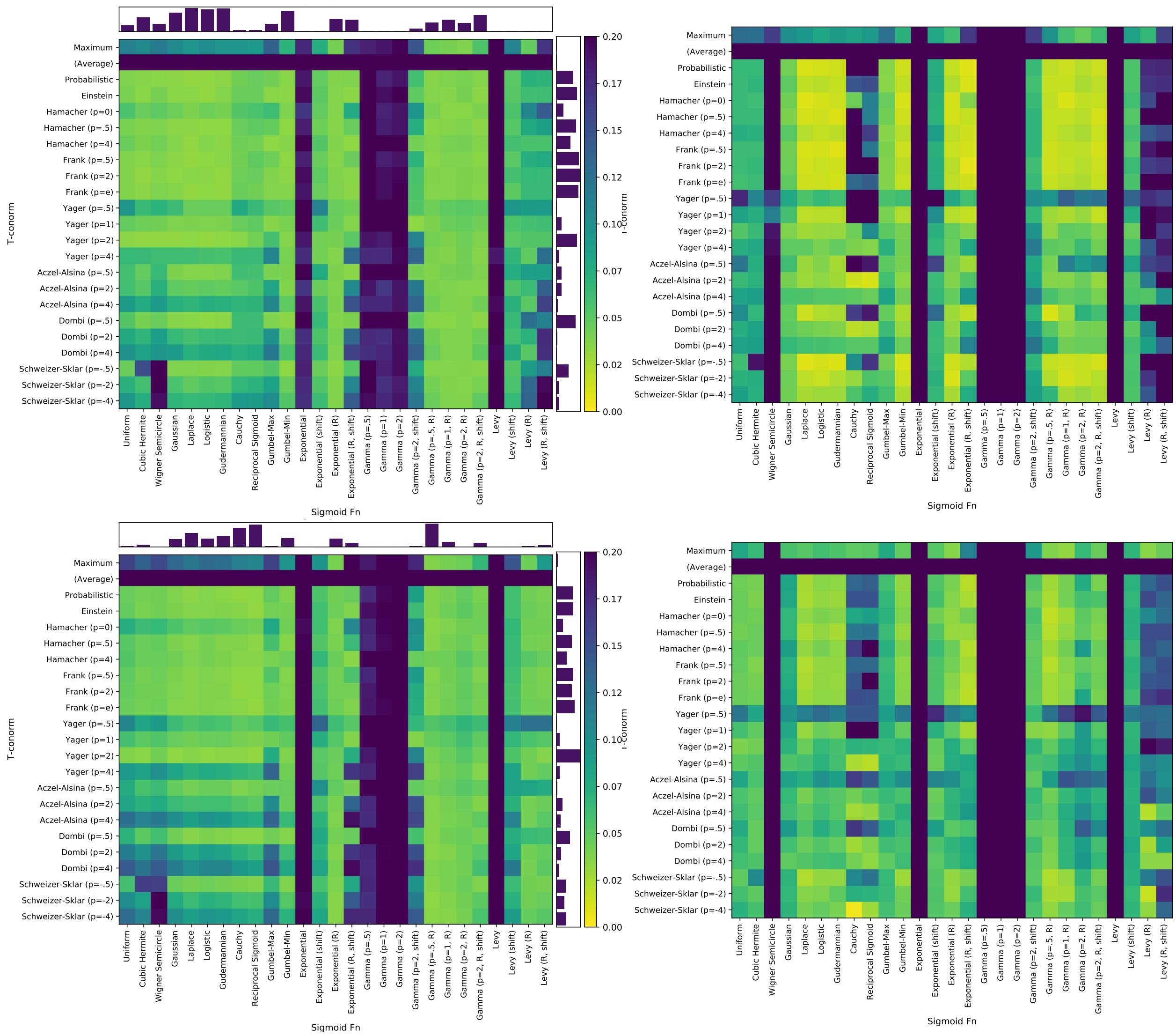

Figure 10. Shape optimization (left) and camera pose optimization (right) applied to a model of a chair. Top: set of original distributions F. Bottom: set of the respective square-root distributions F_{sq}