Generalized Characteristic Function Loss for Crowd Analysis in the Frequency Domain

Supplementary Material

Weibo Shu[1], Jia Wan[1,2], and Antoni B. Chan[1]
[1] Dept. of Computer Science, City University of Hong Kong
[2] Dept. of Electrical and Computer Engineering, University of California

A. Proofs

In this section, we prove the properties/propositions in the paper. We try to keep the proofs in order of their appearance in the main paper, but due to their dependence, there will be some small adjustments.

A.1. Notations and Notes

A.1.1 Notations

- dx, dt, $d\mathbf{x}$, $d\mathbf{t}$ mean the Lebesgue measure.

A.1.2 Notes

Some general knowledge used in the proofs.

- set theory: the cardinality of set, countable set, uncountable set, countable cartesian product.
- measure theory: measure (σ-additivity and continuity of measures), σ-algebra, generated σ-algebra, Dynkin's π-λ theorem (π-system and λ-system).
- real & complex analysis: open set in $\mathbb{R}^n$, Borel sets in $\mathbb{R}^n$, Cauchy-Schwarz inequality, the structure of rational and real numbers, calculus, roots of unity, Euler's formula.
- integral: linearity of integrals to measures, Fubini theorem, dominated convergence theorem, monotone convergence theorem.
- probability: Gaussian distribution, expectation.

A.2. Proofs of properties

Here we provide proofs for Properties 2, 3, 4, and 1.

A.2.1 Proof of Property 2

Proof of Property 2 : Note that $|e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle}| = 1$ and the density map is a finite measure, hence

$$\int_{\mathbb{R}^n} e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} dm(\mathbf{x}) \tag{1}$$

is always integrable for any density map m. Then this property can be easily derived by the linearity of the integral to the finite measure. See the following derivation.

$$\begin{aligned}
\varphi_{m_3}(\mathbf{t}) &= \int_{R^2} e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} dm_3(\mathbf{x}) &(2)\\
&= \int_{R^2} e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} d[(\alpha m_1 + \beta m_2)(\mathbf{x})] &(3)\\
&= \alpha \int_{R^2} e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} dm_1(\mathbf{x}) + \beta \int_{R^2} e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} dm_2(\mathbf{x}) &(4)\\
&= \alpha\varphi_{m_1}(\mathbf{t}) + \beta\varphi_{m_2}(\mathbf{t}). &(5)
\end{aligned}$$

■

A.2.2 Proof of Property 3

Before proving Property 3, we first prove two helpful lemmas.

Lemma 1

$$\lim_{T\to+\infty}\int_0^T \frac{\sin x}{x}\,dx = \frac{\pi}{2}. \tag{6}$$

Proof of Lemma 1 : Suppose $y > 0$, Use integral by parts twice, we have

$$\begin{aligned}
I_y(T) &= \int_0^T e^{-xy}\sin x\,dx &(7)\\
&= (-e^{-xy}\cos x)\big|_0^T - y\int_0^T e^{-xy}\cos x\,dx &(8)\\
&= (-e^{-xy}\cos x)\big|_0^T - y(e^{-xy}\sin x\big|_0^T + y\int_0^T e^{-xy}\sin x\,dx) &(9)\\
&= (-e^{-xy}\cos x - ye^{-xy}\sin x)\big|_0^T - y^2 I_y(T). &(10)
\end{aligned}$$

Hence, we have

$$I_y(T) = \frac{-e^{-Ty}\cos T - ye^{-Ty}\sin T + 1}{1+y^2}. \tag{11}$$

Taking the limit, we have

$$\lim_{T\to+\infty} I_y(T) = \lim_{T\to+\infty}\frac{-e^{-Ty}\cos T - ye^{-Ty}\sin T + 1}{1+y^2} = \frac{1}{1+y^2}. \tag{12}$$

On the domain $(0,T)\times(0,+\infty)$ we have

$$\int_{(0,T)\times(0,+\infty)} |e^{-xy}\sin x| \le \int_{(0,T)\times(0,+\infty)} e^{-xy} = \int_0^T\int_0^{+\infty} e^{-xy}\,dydx = \int_0^T \frac{1}{x}dx < +\infty. \tag{13}$$

Note that the first equality holds by Fubini theorem (e^{-xy} is always positive, which satisfies the condition of using Fubini theorem). Eq. 13 shows that $e^{-xy}\sin x$ is integrable on $(0,T)\times(0,+\infty)$, therefore Fubini theorem can be used to the integral of $e^{-xy}\sin x$ on domain $(0,+\infty)\times(0,T)$. Then by Fubini theorem we have

$$\int_0^T \frac{\sin x}{x} = \int_0^T\int_0^{+\infty} e^{-xy}\sin x\,dydx = \int_0^{+\infty}\int_0^T e^{-xy}\sin x\,dxdy = \int_0^{+\infty} I_y(T)\,dy. \tag{14}$$

Taking the limit and using dominated convergence theorem to have

$$\lim_{T\to+\infty}\int_0^T \frac{\sin x}{x} = \lim_{T\to+\infty}\int_0^{+\infty} I_y(T)\,dy = \int_0^{+\infty}\lim_{T\to+\infty} I_y(T)\,dy = \int_0^{+\infty}\frac{1}{1+y^2}\,dy = \arctan y\big|_0^{+\infty} = \frac{\pi}{2}. \tag{15}$$

■

Lemma 2 (one-dimension inversion formula) *For a finite measure μ defined on the measure space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, suppose there is an interval $[a, b]$ where $a < b$ and $\mu(\{a\}) = \mu(\{b\}) = 0$, then we have*

$$\mu((a,b)) = \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{[-T,T]} \int_a^b \varphi_\mu(t) e^{-\mathrm{i}tx}\, dxdt, \tag{16}$$

where $\varphi_\mu(t)$ means the characteristic function of μ.

Proof of Lemma 2 : This result is a typical result for the characteristic function of the distribution. To prove it for the finite measure, all we need are to ensure that each part in the original proof hold for the finite measure. Now we begin with

$$\lim_{T\to\infty} \frac{1}{(2\pi)} \int_{[-T,T]} \int_a^b \varphi_\mu(t) e^{-\mathrm{i}tx}\, dxdt = \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{[-T,T]} \int_{\mathbb{R}} e^{\mathrm{i}ty} d\mu(y) \int_a^b e^{-\mathrm{i}tx}\, dxdt. \tag{17}$$

Since μ is a finite measure, $[-T, T] \times [a, b]$ has finite Lebesgue measure, and $|e^{\mathrm{i}t(y-x)}| = 1$, then $e^{\mathrm{i}t(y-x)}$ is absolutely integrable on range $\{(x, y, t)|(x, y, t) \in [a, b] \times \mathbb{R} \times [-T, T]\}$ w.r.t. the product measure $\mu(y) \times dx \times dt$, and consequently integrable. Therefore Fubini theorem can be used, and we have

$$\lim_{T\to\infty} \frac{1}{(2\pi)} \int_{[-T,T]} \int_{\mathbb{R}} e^{\mathrm{i}ty} d\mu(y) \int_a^b e^{-\mathrm{i}tx}\, dxdt \tag{18}$$

$$= \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{\mathbb{R}} \int_{[-T,T]} \int_a^b e^{\mathrm{i}t(y-x)}\, dxdtd\mu(y) \tag{19}$$

$$= \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{\mathbb{R}} \int_{[-T,T]} \frac{e^{\mathrm{i}t(y-a)} - e^{\mathrm{i}t(y-b)}}{\mathrm{i}t}\, dtd\mu(y) \tag{20}$$

$$= \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{\mathbb{R}} \left(\int_{[-T,T]} \frac{\cos(t(y-a)) - \cos(t(y-b))}{\mathrm{i}t}\, dt + \int_{[-T,T]} \frac{\mathrm{i}\sin(t(y-a)) - \mathrm{i}\sin(t(y-b))}{\mathrm{i}t}\, dt \right) d\mu(y) \tag{21}$$

$$= \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{\mathbb{R}} \int_{[-T,T]} \frac{\sin(t(y-a)) - \sin(t(y-b))}{t}\, dtd\mu(y). \tag{22}$$

Note that the (22) holds because the first part in the parentheses is 0 by the oddity of the integrand. By integration by substitution and Lemma 1, we have

$$I(y) = \lim_{T\to\infty} \int_{[-T,T]} \frac{\sin(t(y-a)) - \sin(t(y-b))}{t}\, dt \tag{23}$$

$$= \lim_{T\to\infty} \left[\int_{[-T,T]} (y-a) \frac{\sin(t(y-a))}{t(y-a)}\, dt - \int_{[-T,T]} (y-b) \frac{\sin(t(y-b))}{t(y-b)}\, dt \right] \tag{24}$$

$$= \lim_{T\to\infty} \left[2 \int_0^{T(y-a)} \frac{\sin z}{z}\, dz - 2 \int_0^{T(y-b)} \frac{\sin z}{z}\, dz \right] \tag{25}$$

$$= 2 \left[\lim_{T\to\infty} \int_0^{T(y-a)} \frac{\sin z}{z}\, dz - \lim_{T\to\infty} \int_0^{T(y-b)} \frac{\sin z}{z}\, dz \right] \tag{26}$$

$$= \begin{cases} 0 & y < a \\ \pi & y = a \\ 2\pi & a < y < b \\ \pi & y = b \\ 0 & y > b \end{cases}. \tag{27}$$

On the other hand,

$$\left|\int_{[-T,T]} \frac{\sin(t(y-a)) - \sin(t(y-b))}{t}\, dt\right| \le \left|\int_{[-T,T]} \frac{\sin(t(y-a))}{t}\, dt\right| + \left|\int_{[-T,T]} \frac{\sin(t(y-b))}{t}\, dt\right| \tag{28}$$

$$= 2\left|\int_0^{T(y-a)} \frac{\sin z}{z}\, dz\right| + 2\left|\int_0^{T(y-b)} \frac{\sin z}{z}\, dz\right| \tag{29}$$

$$\le 4\sup_w \left|\int_0^w \frac{\sin z}{z}\, dz\right|. \tag{30}$$

By Lemma 1, we have $|\int_0^w \frac{\sin z}{z}\, dz - \frac{\pi}{2}| < \epsilon$ for any $w > P$ where ϵ, P are some positive numbers. Thus, this means $|\int_0^w \frac{\sin z}{z}\, dz| < \epsilon + \frac{\pi}{2}$ if $w > P$. When $w \le P$, $|\int_0^w \frac{\sin z}{z}\, dz| \le \int_0^P |\frac{\sin z}{z}|\, dz \le \int_0^P 1\, dz = P$. Finally we get $\sup_w \left|\int_0^w \frac{\sin z}{z}\, dz\right| \le \max\{\frac{\pi}{2} + \epsilon, P\} = C$. Therefore, we have

$$\left|\int_{[-T,T]} \frac{\sin(t(y-a)) - \sin(t(y-b))}{t}\, dt\right| \le 4C \text{ for any } T > 0. \tag{31}$$

The constant function $4C$ is also integrable on domain $\mathbb{R}$ w.r.t. the finite measure $\mu(y)$. Hence, dominated convergence theorem can be used on (22) to obtain

$$\lim_{T\to\infty} \frac{1}{(2\pi)} \int_{[-T,T]} \int_a^b \varphi_\mu(t) e^{-\mathrm{i}tx}\, dxdt = \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{[-T,T]} \int_{\mathbb{R}} e^{\mathrm{i}ty} d\mu(y) \int_a^b e^{-\mathrm{i}tx}\, dxdt \tag{32}$$

$$= \lim_{T\to\infty} \frac{1}{(2\pi)} \int_{\mathbb{R}} \int_{[-T,T]} \frac{\sin(t(y-a)) - \sin(t(y-b))}{t}\, dtd\mu(y) \tag{33}$$

$$= \frac{1}{(2\pi)} \int_{\mathbb{R}} \lim_{T\to\infty} \left[\int_{[-T,T]} \frac{\sin(t(y-a)) - \sin(t(y-b))}{t}\, dt\right] d\mu(y) \tag{34}$$

$$= \frac{1}{(2\pi)} \int_{\mathbb{R}} I(y)\, d\mu(y) \tag{35}$$

$$= \mu((a,b)) + \frac{\mu(\{a\}) + \mu(\{b\})}{2} \tag{36}$$

$$= \mu((a,b)). \tag{37}$$

Eq. 36 holds by substituting (27) into the integral. The final equality holds by the assumption $\mu(\{a\}) = \mu(\{b\}) = 0$ in the Lemma. ■

Next we are ready to prove Property 3.

Proof of Property 3 : Let $int(A) = (a_1, b_1) \times (a_2, b_2)$ which is the interior of A, since $m(\partial A) = 0$, we have $m(A) = m(int(A))$. We also define $out(A) = A^c = \mathbb{R}^2 - [a_1, b_1] \times [a_2, b_2]$. Similar discussions to the proof of Lemma 2 show the

Fubini theorem can be used. Therefore we have

$$\lim_{T\to+\infty}\frac{1}{(2\pi)^2}\int_{[-T,T]^2}\int_A \varphi_m(\mathbf{t})e^{-\mathrm{i}\langle\mathbf{t},\mathbf{x}\rangle}\,d\mathbf{x}d\mathbf{t} \tag{38}$$

$$=\lim_{T\to+\infty}\frac{1}{(2\pi)^2}\int_{[-T,T]^2}\int_A\int_{\mathbb{R}^2} e^{\mathrm{i}\langle\mathbf{t},\mathbf{y}\rangle}\,dm(\mathbf{y})e^{-\mathrm{i}\langle\mathbf{t},\mathbf{x}\rangle}\,d\mathbf{x}d\mathbf{t} \tag{39}$$

$$=\lim_{T\to\infty}\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\int_{[-T,T]^2}\int_A e^{\mathrm{i}\langle\mathbf{t},(\mathbf{y}-\mathbf{x})\rangle}\,d\mathbf{x}d\mathbf{t}dm(\mathbf{y}) \tag{40}$$

$$=\lim_{T\to\infty}\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\int_{[-T,T]^2}\frac{e^{\mathrm{i}t_1(y_1-a_1)}-e^{\mathrm{i}t_1(y_1-b_1)}}{\mathrm{i}t_1}*\frac{e^{\mathrm{i}t_2(y_2-a_2)}-e^{\mathrm{i}t_2(y_2-b_2)}}{\mathrm{i}t_2}\,dt_1dt_2dm(\mathbf{y}) \tag{41}$$

$$=\lim_{T\to\infty}\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\left(\int_{[-T,T]}\frac{e^{\mathrm{i}t_1(y_1-a_1)}-e^{\mathrm{i}t_1(y_1-b_1)}}{\mathrm{i}t_1}\,dt_1\int_{[-T,T]}\frac{e^{\mathrm{i}t_2(y_2-a_2)}-e^{\mathrm{i}t_2(y_2-b_2)}}{\mathrm{i}t_2}\,dt_2\right)dm(\mathbf{y}) \tag{42}$$

$$=\lim_{T\to\infty}\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\left(\int_{[-T,T]}\frac{\sin(t_1(y_1-a_1))-\sin(t_1(y_1-b_1))}{t_1}\,dt_1\int_{[-T,T]}\frac{\sin(t_2(y_2-a_2))-\sin(t_2(y_2-b_2))}{t_2}\,dt_2\right)dm(\mathbf{y})\,. \tag{43}$$

The last equality holds by applying the result in (22) to both the integral w.r.t. t_1 and t_2. Again, similar discussions for the proof of Lemma 2 show that (see Eq. 27)

$$I(\mathbf{y})=\lim_{T\to\infty}\left(\int_{[-T,T]}\frac{\sin(t_1(y_1-a_1))-\sin(t_1(y_1-b_1))}{t_1}\,dt_1\int_{[-T,T]}\frac{\sin(t_2(y_2-a_2))-\sin(t_2(y_2-b_2))}{t_2}\,dt_2\right) \tag{44}$$

$$=\begin{cases}(2\pi)^2 & \mathbf{y}\in int(A)\\ 2\pi^2 & y_1\in\{a_1,b_1\},y_2\in(a_2,b_2)\\ \pi^2 & y_1\in\{a_1,b_1\},y_2\in\{a_2,b_2\}\\ 2\pi^2 & y_2\in\{a_2,b_2\},y_1\in(a_2,b_2)\\ 0 & \mathbf{y}\in out(A)\end{cases}. \tag{45}$$

We see that the result on the boundary of A is complex, that is why we add the assumption $m(\partial A)=0$ to rule out the complex boundary situation. Finally, similar discussions to the proof of Lemma 2 also show that dominated convergence theorem can be used, and resuming from (43) we have

$$\lim_{T\to+\infty}\frac{1}{(2\pi)^2}\int_{[-T,T]^2}\int_A \varphi_m(\mathbf{t})e^{-\mathrm{i}\langle\mathbf{t},\mathbf{x}\rangle}\,d\mathbf{x}d\mathbf{t} \tag{46}$$

$$=\lim_{T\to\infty}\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\left(\int_{[-T,T]}\frac{\sin(t_1(y_1-a_1))-\sin(t_1(y_1-b_1))}{t_1}\,dt_1\int_{[-T,T]}\frac{\sin(t_2(y_2-a_2))-\sin(t_2(y_2-b_2))}{t_2}\,dt_2\right)dm(\mathbf{y}) \tag{47}$$

$$=\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\lim_{T\to\infty}\left(\int_{[-T,T]}\frac{\sin(t_1(y_1-a_1))-\sin(t_1(y_1-b_1))}{t_1}\,dt_1\int_{[-T,T]}\frac{\sin(t_2(y_2-a_2))-\sin(t_2(y_2-b_2))}{t_2}\,dt_2\right)dm(\mathbf{y}) \tag{48}$$

$$=\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}I(\mathbf{y})\,dm(\mathbf{y}) \tag{49}$$

$$=m(int(A))=m(A). \tag{50}$$

■

Remark of Lemma 1, Lemma 2, Property 3, and chf loss : Note that in Lemma 1, Lemma 2, and Property 3 we use a limit formula in the integral, i.e., we take $\lim_{T\to+\infty}$ rather than the direct Lebesgue integral on the whole integral space $\mathbb{R}$ or $\mathbb{R}^2$. That is because those direct integrals on the whole space do not exist. A function f is Lebesgue integrable means that the positive part f^+'s integral is finite and the negative part f^-'s integral is finite, and then $\int f = \int f^+ - \int f^-$. If f is integrable, then the limit formula integral coincides with the direct integral. But in our case, those functions are actually not integrable on the whole space. For instances, in Lemma 1, $\frac{\sin x}{x}$ is not Lebesgue integrable on $[0, +\infty)$, and $\lim_{T\to+\infty}\int_0^T \frac{\sin x}{x}\, dx$ exists but $\int_0^{+\infty} \frac{\sin x}{x}\, dx$ does not exist. This is also the case in Lemma 2 and Property 3.

But if the function f is non-negative or non-positive, the limit formula integral and the direct Lebesgue integral are always the same by the monotone convergence theorem. And this is the case that we define our chf loss, i.e.,

$$l_{\text{chf}}(m_g, m_p) = \int_{\mathbb{R}^2} |\varphi_{m_g}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t})| d\mathbf{t} = \lim_{T\to+\infty} \int_{[-T,T]^2} |\varphi_{m_g}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t})| d\mathbf{t}. \tag{51}$$

A.2.3 Proof of Property 4

We first prove one lemma before proving Property 4.

Lemma 3 *If m is a discrete density map or a density map by convolving a discrete density map with some Gaussian kernel, then $\int_{\mathbb{R}^2} \|\mathbf{x}\|_2\, dm(\mathbf{x})$ is finite.*

Proof of Lemma 3 : According to the definition of the discrete density map (see Definition 2), if m is a discrete density map, suppose $\{\mathbf{x}_i\}_{i=1}^n$ is all the singletons where m distributes its measure on, then

$$\int_{\mathbb{R}^2} \|\mathbf{x}\|_2\, dm(\mathbf{x}) = \sum_{i=1}^n m(\{\mathbf{x}_i\})\|\mathbf{x_i}\|_2 < +\infty. \tag{52}$$

If $\mathbf{X} = (X_1, X_2) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\begin{aligned}
\mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[\|\mathbf{X}\|_2] &= \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[\mathbb{1}_{\|\mathbf{X}\|_2\le 1}\|\mathbf{X}\|_2] + \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[\mathbb{1}_{\|\mathbf{X}\|_2>1}\|\mathbf{X}\|_2] && (53)\\
&\le \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[\mathbb{1}_{\|\mathbf{X}\|_2\le 1}] + \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[\|\mathbf{X}\|_2^2] && (54)\\
&= \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[\mathbb{1}_{\|\mathbf{X}\|_2\le 1}] + \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[X_1^2] + \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[X_2^2] && (55)\\
&= \mathbb{E}_{\mathbf{X}\sim\mathcal{N}}[\mathbb{1}_{\|\mathbf{X}\|_2\le 1}] + \boldsymbol{\mu}_1^2 + \boldsymbol{\Sigma}_{11} + \boldsymbol{\mu}_2^2 + \boldsymbol{\Sigma}_{22} && (56)\\
&< +\infty. && (57)
\end{aligned}$$

Therefore, if $m = \sum_{i=1}^n \tilde{m}(\boldsymbol{\mu}^{(i)})\mathcal{N}(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma}^{(i)})$, i.e., the density map obtained by convolving a discrete density map with some Gaussian kernel, then by the linearity of integral to finite measures, we have

$$\int_{\mathbb{R}^2} \|\mathbf{x}\|_2\, dm(\mathbf{x}) = \sum_{i=1}^n \tilde{m}(\boldsymbol{\mu}^{(i)})\mathbb{E}_{\mathbf{X}\sim\mathcal{N}(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}^{(i)})}[\|\mathbf{X}\|_2] < +\infty. \tag{58}$$

■

Next we prove Property 4.

Proof of Property 4 : Let $\mathbf{h}$ be any non-zero two-dimensional vector in $\mathbb{R}^2$, φ be the characteristic function, then

$$\begin{aligned}
|\varphi_m(\mathbf{t}+\mathbf{h}) - \varphi_m(\mathbf{t})| &= \left|\int_{\mathbb{R}^2} (e^{\mathrm{i}\langle \mathbf{t}+\mathbf{h},\mathbf{x}\rangle} - e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle})\, dm(\mathbf{x})\right| && (59)\\
&\le \int_{\mathbb{R}^2} \left|e^{\mathrm{i}\langle \mathbf{t}+\mathbf{h},\mathbf{x}\rangle} - e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle}\right| dm(\mathbf{x}) && (60)\\
&= \int_{\mathbb{R}^2} |e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle}||e^{\mathrm{i}\langle \mathbf{h},\mathbf{x}\rangle} - 1|\, dm(\mathbf{x}) = \int_{\mathbb{R}^2} |e^{\mathrm{i}\langle \mathbf{h},\mathbf{x}\rangle} - 1|\, dm(\mathbf{x}) && (61)\\
&\le \int_{\mathbb{R}^2} |\langle \mathbf{h},\mathbf{x}\rangle|\, dm(\mathbf{x}) && (62)\\
&\le \int_{\mathbb{R}^2} \|\mathbf{h}\|_2\|\mathbf{x}\|_2\, dm(\mathbf{x}) = \|\mathbf{h}\|_2 \int_{\mathbb{R}^2} \|\mathbf{x}\|_2\, dm(\mathbf{x}) = C\|\mathbf{h}\|_2, && (63)
\end{aligned}$$

where $C = \int_{\mathbb{R}^2} \|\mathbf{x}\|_2 \, dm(\mathbf{x}) < +\infty$ by Lemma 3. Eq. 60 holds by the absolute integral inequality; Eq. 61 holds by $|e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle}| = 1$; Eq. 62 holds by the property $|e^{\mathrm{i}x} - 1| \leq |x|$; Eq. 63 holds by the Cauchy–Schwarz inequality.

More generally, if m is a finite measure defined on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ where $\mathcal{B}_{\mathbb{R}^n}$ is all the Borel sets, then its characteristic function φ_m is uniformly continuous. This is because $|e^{\mathrm{i}\langle \mathbf{h},\mathbf{x}\rangle} - 1| \leq 2$, and m is a finite measure means $\int_{\mathbb{R}^2} 2 \, dm(\mathbf{x})$ is finite, then dominated convergence theorem can be used to obtain

$$\lim_{\mathbf{h}\to 0} \int_{\mathbb{R}^2} |e^{\mathrm{i}\langle \mathbf{h},\mathbf{x}\rangle} - 1| \, dm(\mathbf{x}) = \int_{\mathbb{R}^2} \lim_{\mathbf{h}\to 0} |e^{\mathrm{i}\langle \mathbf{h},\mathbf{x}\rangle} - 1| \, dm(\mathbf{x}) = 0. \tag{64}$$

Combining with (63), we have

$$|\varphi_m(\mathbf{t}+\mathbf{h}) - \varphi_m(\mathbf{t})| \leq \int_{\mathbb{R}^2} |e^{\mathrm{i}\langle \mathbf{h},\mathbf{x}\rangle} - 1| \, dm(\mathbf{x}) \longrightarrow 0 \quad \text{as} \quad \mathbf{h} \to 0. \tag{65}$$

This shows that φ_m is uniformly continuous.

∎

A.2.4 Proof of Property 1

We first prove 3 lemmas before proving Property 1.

Lemma 4 *Suppose μ is a finite measure defined on n-dimension measure space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ where $\mathcal{B}_{\mathbb{R}^n}$ means all the Borel sets, then there are only countable slices with non-zero measures w.r.t. μ, i.e.,*

$$V_i = \{c_i | \mu(\mathbb{R}^{i-1} \times \{c_i\} \times \mathbb{R}^{n-i}) \neq 0\} \tag{66}$$

is a countable set for $1 \leq i \leq n$. Here, $\mathbb{R}^0$ means not taking any cartesian product, i.e., $\mathbb{R}^0 \times A = A \times \mathbb{R}^0 = A$.

Proof of Lemma 4 : We prove this lemma by contradiction for $i = 1$. The other cases are the same. Since

$$V_1 = \{c_1 | \mu(\{c_1\} \times \mathbb{R}^{n-1}) \neq 0\} = \bigcup_{k=1}^{+\infty} \{c_1 | \mu(\{c_1\} \times \mathbb{R}^{n-1}) > \frac{1}{k}\}, \tag{67}$$

then if V_1 is not countable, there must be some $\tilde{k}$ s.t. $V_1(\tilde{k}) = \{c_1 | \mu(\{c_1\} \times \mathbb{R}^{n-1}) > \frac{1}{\tilde{k}}\}$ has infinite elements – the cardinality of the union of countably many finite sets is countably infinite, and it can not go up to uncountably infinite; Hence there must be some infinite sets. We take countably many elements from $V_1(\tilde{k})$ and use the σ-additivity to obtain

$$\mu(\bigcup_{j=1}^{+\infty} \{\{c_1^j\} \times \mathbb{R}^{n-1}\}) = \sum_{j=1}^{+\infty} \mu(\{c_1^j\} \times \mathbb{R}^{n-1}) \geq +\infty * \frac{1}{\tilde{k}} = +\infty, \tag{68}$$

which contradicts the truth μ is a finite measure. ∎

Lemma 5 *Suppose $A \subset \mathbb{R}$ is a countable set, then there is a countable subset $B \subset \mathbb{R} - A$ s.t. $\{x | a < x < b \ \wedge \ x \in B\}$ is an infinite set for any $a < b$ and $a, b \in \mathbb{R}$.*

Proof of Lemma 5 : For a real number r, we define set $\mathbb{Q}(r) = \mathbb{Q} + r$ where $\mathbb{Q}$ is all the rational numbers. Then, for two different real numbers r_1 and r_2, there are only two possibilities:

1. $\mathbb{Q}(r_1) = \mathbb{Q}(r_2)$;
2. $\mathbb{Q}(r_1) \cap \mathbb{Q}(r_2) = \emptyset$.

Suppose $\mathbb{Q}(r_1) \cap \mathbb{Q}(r_2) \neq \emptyset$, then there exists rational numbers q_1 and q_2 s.t. $q_1 + r_1 = q_2 + r_2$, which means $r_2 - r_1 = q_1 - q_2 = q_3$ is a rational number. Then for any element $q + r_2 \in \mathbb{Q}(r_2)$, we can find $q + q_3 + r_1 \in \mathbb{Q}(r_1)$ that equals $q + r_2$. Hence, $\mathbb{Q}(r_2) \subset \mathbb{Q}(r_1)$, and similar discussions lead to $\mathbb{Q}(r_1) \subset \mathbb{Q}(r_2)$. Therefore, either $\mathbb{Q}(r_1) = \mathbb{Q}(r_2)$ or $\mathbb{Q}(r_1) \cap \mathbb{Q}(r_2) = \emptyset$.

Since $\mathbb{Q}(r)$ is a countable set for any $r \in \mathbb{R}$ and we have

$$\mathbb{R} = \bigcup_{s \in \{\mathbb{Q}(r) | r \in \mathbb{R}\}} s, \tag{69}$$

then $\{\mathbb{Q}(r) | r \in \mathbb{R}\}$ is uncountably infinite. Otherwise, $\mathbb{R}$ is the union of countably many countable sets, which means $\mathbb{R}$ is a countable set, contradicting the truth $\mathbb{R}$ is an uncountable set. On the other hand, A is countable subset of $\mathbb{R}$, and we have proved that any pair in $\{\mathbb{Q}(r) | r \in \mathbb{R}\}$ has empty intersection, therefore there must be some real number r^* satisfying $A \cap \mathbb{Q}(r^*) = \emptyset$. Then $B = \mathbb{Q}(r^*)$ is the desired set satisfying all the conditions in the lemma. ■

Lemma 6 *Suppose μ_1 and μ_2 are two finite measures defined on n-dimension measure space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ where $\mathcal{B}_{\mathbb{R}^n}$ means all the Borel sets, then there is a π-system $\mathcal{S}$ who is a countable set satisfying*

- $\forall A \in \mathcal{S}$, $A = \prod_{i=1}^{n} (a_i, b_i)$ *and* $\mu_1(\partial A) = \mu_2(\partial A) = 0$,
- $\sigma(\mathcal{S}) = \mathcal{B}_{\mathbb{R}^n}$,

where ∂A means the boundary of A and $\sigma(\mathcal{S})$ means the generated σ-algebra of $\mathcal{S}$.

Proof of Lemma 6 : By Lemma 4, we know $V_i^1 = \{c_i | \mu_1(\mathbb{R}^{i-1} \times \{c_i\} \times \mathbb{R}^{n-i}) \neq 0\}$ and $V_i^2 = \{c_i | \mu_2(\mathbb{R}^{i-1} \times \{c_i\} \times \mathbb{R}^{n-i}) \neq 0\}$ are both countable sets for $1 \leq i \leq n$, then $V_i = V_i^1 \cup V_i^2$ is countable set for $1 \leq i \leq n$. Therefore, by Lemma 5, it is not hard to find a countable set $P_i \subset \mathbb{R}$ with

1. $\{x | a < x < b \ \wedge \ x \in P_i\}$ is an infinite set for any $a < b$ and $a, b \in \mathbb{R}$;
2. $\forall x \in P_i$, $\mu_1(\mathbb{R}^{i-1} \times \{x_i = x\} \times \mathbb{R}^{n-i}) = \mu_2(\mathbb{R}^{i-1} \times \{x_i = x\} \times \mathbb{R}^{n-i}) = 0$

for each i in $1 \leq i \leq n$. Let

$$\mathcal{S} = \{\prod_{i=1}^{n} (a_i, b_i) | a_i, b_i \in P_i \ for \ 1 \leq i \leq n\} \tag{70}$$

be the set of some special open boxes, then we claim $\mathcal{S}$ is the desired family of sets. Firstly, since P_i is countable for all $1 \leq i \leq n$, the cardinality of $\mathcal{S}$ is the cardinality of the cartesian product of finite countable sets, which is countable. Secondly, we have

$$\prod_{i=1}^{n} (a_i, b_i) \cap \prod_{i=1}^{n} (c_i, d_i) = \prod_{i=1}^{n} (a_i, b_i) \cap (c_i, d_i). \tag{71}$$

Then, the intersection of two open boxes in $\mathcal{S}$ still falls into $\mathcal{S}$, which shows $\mathcal{S}$ is a π-system. Thirdly, by the second property of P_i, we have $\forall A \in \mathcal{S}$, $\mu_1(\partial A) = \mu_2(\partial A) = 0$. Fourthly, suppose U is non-empty open set, then for each $x \in U$, we can find an open ball $B(x, r) \subset U$. Further we can find an open box $A[x] \in \mathcal{S}$ such that $x \in A \subset B(x, r) \subset U$ by the first property of P_i. Then $U = \bigcup_{x \in U} A[x]$. But $A[x] \in \mathcal{S}$ and $\mathcal{S}$ is a countable set, it shows U is actually the union of countably many $A[x]$. Then, any open set can be represented by the countable union of the sets in $\mathcal{S}$. Define $\mathcal{T}$ as the family of all the open sets in $\mathbb{R}^n$, then the result shows

$$\sigma(\mathcal{S}) \supset \sigma(\mathcal{T}) = \mathcal{B}_{\mathbb{R}^n}. \tag{72}$$

On the other hand, all the sets in $\mathcal{S}$ are originally open sets, which means

$$\sigma(\mathcal{S}) \subset \sigma(\mathcal{T}) = \mathcal{B}_{\mathbb{R}^n}. \tag{73}$$

Finally we get $\sigma(\mathcal{S}) = \mathcal{B}_{\mathbb{R}^n}$. ■

Now, we can prove Property 1.

Proof of Property 1 : If $m_1 = m_2$, it is obvious that $\varphi_{m_1} = \varphi_{m_2}$ according the definition of characteristic functions (see Definition 4). We now prove the other side.

Suppose $\varphi_{m_1} = \varphi_{m_2}$ $a.e.$, then by the continuity of characteristic functions (see Remark of Property 4 in the supplementary material), $\varphi_{m_1} = \varphi_{m_2}$. By Lemma 6, we can find a family of sets $\mathcal{S}$ satisfying:

1. $\mathcal{S}$ is a countable set;
2. $\mathcal{S}$ is a π-system;
3. $\forall A \in \mathcal{S}$, $A = \prod_{i=1}^{2}(a_i, b_i)$ and $m_1(\partial A) = m_2(\partial A) = 0$;
4. $\sigma(\mathcal{S}) = \mathcal{B}_{\mathbb{R}^2}$.

According to the 3rd attribute of $\mathcal{S}$ and Property 3 (the inversion formula), we have

$$\forall A \in \mathcal{S},\ m_1(A) = m_2(A). \tag{74}$$

Now, we define the family of sets $\mathcal{R}$ as

$$\mathcal{R} = \{A \in \mathcal{B}_{\mathbb{R}^2} | m_1(A) = m_2(A)\}. \tag{75}$$

Then we prove $\mathcal{R}$ is a λ-system. Firstly, since $\varphi_{m_1} = \varphi_{m_2}$, then $m_1(\mathbb{R}^2) = \varphi_{m_1}(0) = \varphi_{m_2}(0) = m_2(\mathbb{R}^2)$, which shows the total space $\mathbb{R}^2 \in \mathcal{R}$. Secondly, if $A, B \in \mathcal{R}$ and $A \subset B$, then $m_1(B - A) = m_1(B) - m_1(A) = m_2(B) - m_2(A) = m_2(B - A)$ by σ-additivity of measures. Therefore $B - A \in \mathcal{R}$. Thirdly, if $\{A_i\}_{i=1}^{+\infty} \subset \mathcal{R}$ and $A_i \subset A_{i+1}$ for $i \geq 1$, then by the continuity of measures we have $m_1(\bigcup_{i=1}^{+\infty} A_i) = \lim_{i\to+\infty} m_1(A_i) = \lim_{i\to+\infty} m_2(A_i) = m_2(\bigcup_{i=1}^{+\infty} A_i)$, which means $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{R}$. In summary, $\mathcal{S}$ is a λ-system.

Since $\mathcal{S} \subset \mathcal{R}$ and $\mathcal{S}$ is a π-system (the 2nd attribute of $\mathcal{S}$), then by Dynkin's π-λ theorem, we have $\sigma(\mathcal{S}) \subset \mathcal{R}$. By the 4th attribute of $\mathcal{S}$, it means $\mathcal{B}_{\mathbb{R}^2} \subset \mathcal{R}$. But according to the definition of $\mathcal{R}$, $\mathcal{R} \subset \mathcal{B}_{\mathbb{R}^2}$. Therefore, we have $\mathcal{B}_{\mathbb{R}^2} = \mathcal{R}$, which means $m_1(A) = m_2(A)$ for any $A \in \mathcal{B}_{\mathbb{R}^2}$. ∎

A.3. Proofs of Propositions

A.3.1 Proof of Proposition 1

Proof of Proposition 1 : $l_{chf}(\hat{m}_g, m_p) = 0$ means

$$\int_{\mathbb{R}^2} |\varphi_{\hat{m}_g}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t})| d\mathbf{t} = 0. \tag{76}$$

Let

$$f(\mathbf{t}) = |\varphi_{\hat{m}_g}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t})| \geq 0, \tag{77}$$

and $\{r_i\}_{i=1}^{+\infty}$ is a series of decreasing positive real numbers converging to 0, then we have

$$\{\mathbf{t} \in \mathbb{R}^2 | f(\mathbf{t}) > 0\} = \bigcup_{i=1}^{+\infty} \{\mathbf{t} \in \mathbb{R}^2 | f(\mathbf{t}) > r_i\}. \tag{78}$$

If $\mathcal{L}(\{f(\mathbf{t}) > r_i\}) \neq 0$, where $\mathcal{L}$ is the Lebesgue measure, then

$$0 < \mathcal{L}(\{f(\mathbf{t}) > r_i\}) * r_i = \int_{f(\mathbf{t})>r_i} r_i d\mathbf{t} \tag{79}$$

$$\leq \int_{f(\mathbf{t})>r_i} f(\mathbf{t}) d\mathbf{t} \leq \int_{\mathbb{R}^2} f(\mathbf{t}) d\mathbf{t} = 0, \tag{80}$$

which is a contradiction. Hence $\mathcal{L}(\{f(\mathbf{t}) > r_i\}) = 0$ for all i, then by the continuity of the Lebesgue measure,

$$\mathcal{L}(\{f(\mathbf{t}) > 0\}) = \mathcal{L}(\bigcup_{i=1}^{+\infty} \{f(\mathbf{t}) > r_i\}) = \lim_{i=1}^{+\infty} \mathcal{L}(\{f(\mathbf{t}) > r_i\}) = 0. \tag{81}$$

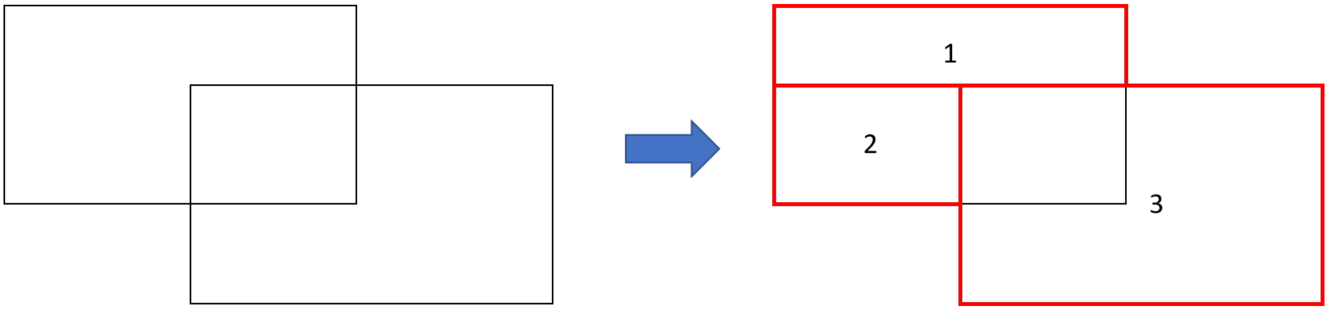


Figure 1. The union of two boxes can be written as the union of disjoint boxes with vertex coordinates in the same vertex coordinate set of the original two boxes.

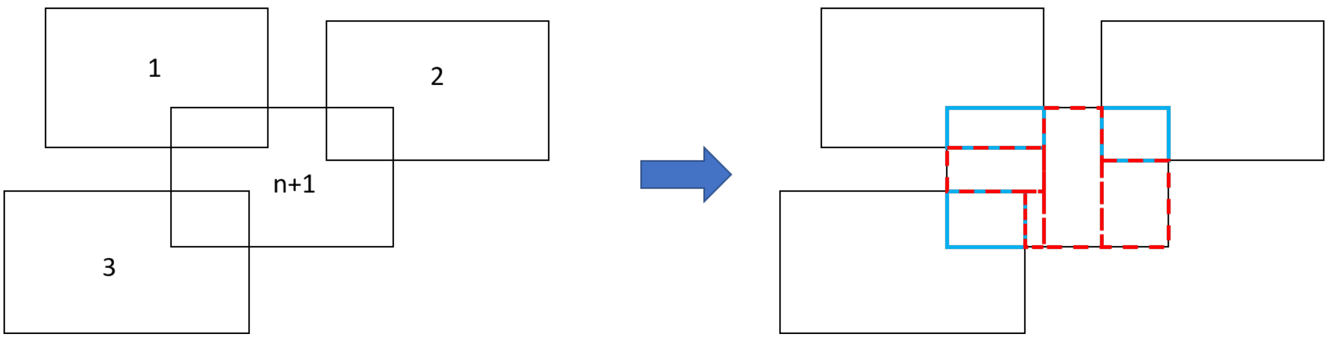


Figure 2. The incremental split. Suppose 1, 2, and 3 are the split disjoint boxes before, $n+1$ is the new coming box. The method first removes the intersection between $n+1$ (in this example $n=3$, for clearness we use $n+1$ rather than 4 to mark the new coming box) and k ($k=1,2,3$) from $n+1$ (the blue boxes in the right pattern). Second, the method splits the remaining part of $n+1$ into disjoint boxes (the red boxes in the right pattern). The strategy can always work for finite n.

This means

$$\varphi_{\hat{m}_g}(\mathbf{t}) = \varphi_{m_p}(\mathbf{t}) \quad a.e. \tag{82}$$

Then by Property 1, $\hat{m}_g = m_p$.

Hence, we have $l_{chf}(\hat{m}_g, m_p) = 0 \longrightarrow \hat{m}_g = m_p$, which means the chf loss is not underdetermined for $\hat{m}_g$ and m_p ■

A.3.2 Proof of Proposition 2

To prove Proposition 2, we first prove one lemma.

Lemma 7 *For a set A who is the union of countably many open boxes $\bigcup_{i=1}^{+\infty}(a_1^{(i)}, b_1^{(i)}) \times (a_2^{(i)}, b_2^{(i)})$, suppose the first vertex set is $V_1 = \{a_1^{(i)}, b_1^{(i)}\}_{i=1}^{+\infty}$ and the second vertex set is $V_2 = \{a_2^{(i)}, b_2^{(i)}\}_{i=1}^{+\infty}$. Then $A = \bigcup_{j=1}^{+\infty} A_j$ where $A_{j_1} \cap A_{j_2} = \emptyset$ if $j_1 \neq j_2$. And for all j, A_j has the form of boxes: $⟦c_1^{(j)}, d_1^{(j)}⟧ \times ⟦c_2^{(j)}, d_2^{(j)}⟧$, where $⟦$ and $⟧$ mean the boundary can be either open "$(,)$" or closed "$[,]$". According these choices, there are all $2^4 = 16$ types of boxes. Besides, $c_1^{(j)}, d_1^{(j)} \in V_1$ and $c_2^{(j)}, d_2^{(j)} \in V_2$.*

Proof of Lemma 7 : We use proof by induction. If $A = \bigcup_{i=1}^{2}(a_1^{(i)}, b_1^{(i)}) \times (a_2^{(i)}, b_2^{(i)})$, then we can always split A into disjoint boxes satisfying the conditions in this lemma. There are several cases for the union of two boxes, we do not discuss them one by one. The assertion is always true, and Fig. 1 shows one case. Other cases are easily checked. Inductively, if $A = \bigcup_{i=1}^{n}(a_1^{(i)}, b_1^{(i)}) \times (a_2^{(i)}, b_2^{(i)})$ can be split into disjoint boxes with vertexes's coordinates satisfying conditions in the lemma, then for $A = \bigcup_{i=1}^{n+1}(a_1^{(i)}, b_1^{(i)})$, we can first split $\bigcup_{i=1}^{n}(a_1^{(i)}, b_1^{(i)})$ into those boxes, and then do the split for the $(n+1)$-th box as shown in Fig. 2.

The induction shows the lemma holds for all the cases that A is the union of finite boxes. If $A = \bigcup_{i=1}^{+\infty}(a_1^{(i)}, b_1^{(i)}) \times (a_2^{(i)}, b_2^{(i)})$, let $C_k = \bigcup_{i=1}^{k}(a_1^{(i)}, b_1^{(i)}) \times (a_2^{(i)}, b_2^{(i)})$, then $A = \bigcup_{k=1}^{+\infty} C_k$ and $C_k \subset C_{k+1}$. If we use the incremental split method shown in Fig. 2, then we can write each C_k as the union of a family $\mathcal{C}_k$ of disjoint boxes and $\mathcal{C}_k \subset \mathcal{C}_{k+1}$. Then

$$A = \bigcup_{k=1}^{+\infty} \bigcup_{B_i \in \mathcal{C}_k} B_i \ , \ B_i \text{ are boxes.} \tag{83}$$

Since $\mathcal{C}_k \subset \mathcal{C}_{k+1}$ and each $\mathcal{C}_k$ is a finite family of disjoint boxes whose vertexes satisfy the conditions in the lemma, A is actually the union of countably many disjoint boxes whose vertexes satisfy the conditions in the lemma. This completes the proof. ■

Proof of Proposition 2 : Since $\widehat{m_g}$ and m_p are both finite measures on $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$, by Lemma 6 and its proof, there is a countable family of sets $\mathcal{S}$ satisfying:

1. $\forall A \in \mathcal{S}$, $A = \prod_{i=1}^{2}(a_i, b_i)$ where $\widehat{m_g}(\{a_1\} \times \mathbb{R}) = m_p(\{a_1\} \times \mathbb{R}) = \widehat{m_g}(\{b_1\} \times \mathbb{R}) = m_p(\{b_1\} \times \mathbb{R}) = 0$ and $\widehat{m_g}(\mathbb{R} \times \{a_2\}) = m_p(\mathbb{R} \times \{a_2\}) = \widehat{m_g}(\mathbb{R} \times \{b_2\}) = m_p(\mathbb{R} \times \{b_2\}) = 0$, therefore $\widehat{m_g}(\partial A) = m_p(\partial A) = 0$.
2. for any open set U, $U = \bigcup_{i=1}^{+\infty} A_i$ where $A_i \in \mathcal{S}$ for each i.

Use Property 3 and the 1st attribute of $\mathcal{S}$, for any set $A \in \mathcal{S}$, we can get

$$|\widehat{m_g}(A) - m_p(A)| \tag{84}$$

$$= \left| \lim_{T\to+\infty} \frac{1}{(2\pi)^2} \int_{[-T,T]^2} \int_A \varphi_{\widehat{m_g}}(\mathbf{t}) e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \, d\mathbf{x}d\mathbf{t} - \lim_{T\to+\infty} \frac{1}{(2\pi)^2} \int_{[-T,T]^2} \int_A \varphi_{m_p}(\mathbf{t}) e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \, d\mathbf{x}d\mathbf{t} \right| \tag{85}$$

$$= \left| \lim_{T\to+\infty} \frac{1}{(2\pi)^2} \int_{[-T,T]^2} \int_A [\varphi_{\widehat{m_g}}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t})] e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \, d\mathbf{x}d\mathbf{t} \right| \tag{86}$$

$$= \lim_{T\to+\infty} \left| \frac{1}{(2\pi)^2} \int_{[-T,T]^2} \int_A [\varphi_{\widehat{m_g}}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t})] e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \, d\mathbf{x}d\mathbf{t} \right| \tag{87}$$

$$\le \lim_{T\to+\infty} \frac{1}{(2\pi)^2} \int_{[-T,T]^2} \int_A \left| [\varphi_{\widehat{m_g}}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t})] e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \right| \, d\mathbf{x}d\mathbf{t} \tag{88}$$

$$= \lim_{T\to+\infty} \frac{1}{(2\pi)^2} \int_{[-T,T]^2} \int_A \left| \varphi_{\widehat{m_g}}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t}) \right| \, d\mathbf{x}d\mathbf{t} \tag{89}$$

$$= \lim_{T\to+\infty} \frac{1}{(2\pi)^2} \int_A \int_{[-T,T]^2} \left| \varphi_{\widehat{m_g}}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t}) \right| \, d\mathbf{t}d\mathbf{x} \tag{90}$$

$$= \frac{1}{(2\pi)^2} \int_A \lim_{T\to+\infty} \int_{[-T,T]^2} \left| \varphi_{\widehat{m_g}}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t}) \right| \, d\mathbf{t}d\mathbf{x} \tag{91}$$

$$= \frac{1}{(2\pi)^2} \int_A \int_{\mathbb{R}^2} \left| \varphi_{\widehat{m_g}}(\mathbf{t}) - \varphi_{m_p}(\mathbf{t}) \right| \, d\mathbf{t}d\mathbf{x} \tag{92}$$

$$= (2\pi)^{-2} \int_A l_{chf}(\widehat{m_g}, m_p) \, d\mathbf{x} = (2\pi)^{-2} l_{chf}(\widehat{m_g}, m_p) \mathcal{L}(A). \tag{93}$$

(87) holds by the continuity of the absolute value function; (88) holds by the absolute integral inequality; (89) holds by the truth $|e^{\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle}| = 1$; (90) holds by Fubini theorem (the integrand is non-negative measurable function, hence Fubini theorem can be used); (91) and (92) hold by monotone convergence theorem.

For any open set U, by the 2nd attribute of $\mathcal{S}$, $U = \bigcup_{i=1}^{+\infty} A_i$ where $A_i \in \mathcal{S}$ for each i. Then by Lemma 7 and the first attribute of $\mathcal{S}$, U is actually the union of countably many disjoint boxes B_j whose boundary has zero measure under both $\widehat{m_g}$ and m_p. Suppose $c_1^{(j)}$, $d_1^{(j)}$ are first dimension's coordinates of B_j's vertexes, and $c_2^{(j)}$, $d_2^{(j)}$ are second dimension's coordinates of B_j's vertexes, then the interior of B_j is $int(B_j) = (c_1^{(j)}, d_1^{(j)}) \times (c_2^{(j)}, d_2^{(j)})$. Since it has zero measure boundary under both $\widehat{m_g}$ and m_p, we have $\widehat{m_g}(int(B_j)) = \widehat{m_g}(B_j)$ and $m_p(int(B_j)) = m_p(B_j)$. Finally, for any open set U, we have

$$|\widehat{m_g}(U) - m_p(U)| = \left|\widehat{m_g}(\bigcup_{j=1}^{+\infty} B_j) - m_p(\bigcup_{j=1}^{+\infty} B_j)\right| \tag{94}$$

$$= \left|\sum_{j=1}^{+\infty}\widehat{m_g}(B_j) - \sum_{j=1}^{+\infty} m_p(B_j)\right| \tag{95}$$

$$= \left|\sum_{j=1}^{+\infty}[\widehat{m_g}(B_j) - m_p(B_j)]\right| \tag{96}$$

$$\leq \sum_{j=1}^{+\infty}|\widehat{m_g}(B_j) - m_p(B_j)| = \sum_{j=1}^{+\infty}|\widehat{m_g}(int(B_j)) - m_p(int(B_j))| \tag{97}$$

$$\leq \sum_{j=1}^{+\infty}(2\pi)^{-2} l_{chf}(\widehat{m_g}, m_p)\mathcal{L}(int(B_j)) \tag{98}$$

$$= (2\pi)^{-2} l_{chf}(\widehat{m_g}, m_p)\sum_{j=1}^{+\infty}\mathcal{L}(B_j) \tag{99}$$

$$= (2\pi)^{-2} l_{chf}(\widehat{m_g}, m_p)\mathcal{L}(U). \tag{100}$$

(95) holds by the σ-additivity of measures; (96) holds by the limit law of convergence series; (98) holds by (84-93); (99) holds by the truth that all the boxes have zero-measure boundary under Lebesgue measure. (100) again holds by the σ-additivity of measures.

■

A.3.3 Proof of Proposition 3

Proof of Proposition 3 : For the mean of $\varphi_{\widetilde{m}_g}$ we have

$$\mathbb{E}_{\boldsymbol{\epsilon}_j\sim\mathcal{N}(0,\boldsymbol{\Lambda})}[\varphi_{\widetilde{m}_g}(\mathbf{t})] = \sum_{j=1}^{M}\exp(\mathrm{i}\boldsymbol{\mu}_j^T\mathbf{t})\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T\mathbf{t})]\exp(-\tfrac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}) \tag{101}$$

$$= \sum_{j=1}^{M}\exp(\mathrm{i}\boldsymbol{\mu}_j^T\mathbf{t})\exp(-\tfrac{1}{2}\mathbf{t}^T\boldsymbol{\Lambda}\mathbf{t})\exp(-\tfrac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}) \tag{102}$$

$$= \sum_{j=1}^{M}\exp(\mathrm{i}\boldsymbol{\mu}_j^T\mathbf{t})\exp(-\tfrac{1}{2}\mathbf{t}^T(\boldsymbol{\Lambda}+\boldsymbol{\Sigma})\mathbf{t}), \tag{103}$$

where the expectation in the first line is exactly the characteristic function of $\boldsymbol{\epsilon}_j \sim \mathcal{N}(0, \boldsymbol{\Lambda})$.

Next, we compute the variance of $\varphi_{\widetilde{m}_g}$,

$$\mathrm{var}_{\boldsymbol{\epsilon}_j\sim\mathcal{N}(0,\boldsymbol{\Lambda})}(\varphi_{\widetilde{m}_g}(\mathbf{t})) = \mathrm{var}\Big(\sum_{j=1}^{M}\exp(\mathrm{i}\boldsymbol{\mu}_j^T\mathbf{t})\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T\mathbf{t})\exp(-\tfrac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t})\Big) \tag{104}$$

$$= \sum_{j=1}^{M}\big|\exp(\mathrm{i}\boldsymbol{\mu}_j^T\mathbf{t})\exp(-\tfrac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t})\big|^2\mathrm{var}(\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T\mathbf{t})) \tag{105}$$

$$= \sum_{j=1}^{M}\exp(-\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t})\mathrm{var}(\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T\mathbf{t})), \tag{106}$$

where (105) uses the property that $\text{var}(\sum_j a_j z_j) = \sum_j |a_j|^2 \text{var}(z_j)$ for independent complex r.v.s z_j with coefficients $a_j \in \mathbb{C}$, and (106) uses $|\exp(\mathrm{i}\boldsymbol{\mu}_j^T \mathbf{t})|^2 = 1$. For each variance term,

$$\begin{aligned}\text{var}(\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T \mathbf{t})) &= \mathbb{E}[|\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T \mathbf{t})|^2] - |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T \mathbf{t})]|^2 && (107)\\ &= 1 - \left|\exp(-\tfrac{1}{2}\mathbf{t}^T \boldsymbol{\Lambda} \mathbf{t})\right|^2 && (108)\\ &= 1 - \exp(-\mathbf{t}^T \boldsymbol{\Lambda} \mathbf{t}). && (109)\end{aligned}$$

Finally, substituting into (106), we have

$$\text{var}_{\boldsymbol{\epsilon}_j \sim \mathcal{N}(0,\boldsymbol{\Lambda})}(\varphi_{\widetilde{m}_g}(\mathbf{t})) = M(1 - \exp(-\mathbf{t}^T \boldsymbol{\Lambda} \mathbf{t}))\exp(-\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}). \tag{110}$$

■

A.3.4 Proof of Proposition 4

Proof of Proposition 4 : Note that (106) holds by the i.i.d. assumption and regardless of the specific distribution of the annotation noise $\boldsymbol{\epsilon}_j$. Suppose $\boldsymbol{\epsilon}_j$ are i.i.d., but we do not make any assumption about their distribution, which we refer to as $\boldsymbol{\epsilon}$ for simplicity. Then we have

$$\begin{aligned}\text{var}(\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T \mathbf{t})) &= \mathbb{E}[|\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T \mathbf{t})|^2] - |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}_j^T \mathbf{t})]|^2 && (111)\\ &= 1 - |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})]|^2. && (112)\end{aligned}$$

Substituting into (106), we have

$$\text{var}(\varphi_{\widetilde{m}_g}(\mathbf{t})) = M(1 - |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})]|^2)\exp(-\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}). \tag{113}$$

Thus (113) indicates that the variance still linearly scales with the count M, regardless of the distribution of the annotation noise. Moreover, since $|\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})| = 1$, by the absolute integral inequality, we have

$$0 \leq |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})]|^2 \leq \mathbb{E}[|\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})|]^2 = 1. \tag{114}$$

As a result the term $(1 - |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})]|^2)$ is bounded within $[0, 1]$, and thus we have

$$0 \leq \lim_{|\mathbf{t}|\to\infty} \text{var}(\varphi_{\widetilde{m}_g}(\mathbf{t})) \leq \lim_{|\mathbf{t}|\to\infty} M\exp(-\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}) = 0. \tag{115}$$

Hence we have $\text{var}(\varphi_{\widetilde{m}_g}(\mathbf{t})) \to 0$ when $|\mathbf{t}| \to \infty$. Furthermore, by the dominated convergence theorem, we have

$$\lim_{|\mathbf{t}|\to 0} |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})]|^2 = |\mathbb{E}[\lim_{|\mathbf{t}|\to 0} \exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})]|^2 = |\mathbb{E}[1]|^2 = 1, \tag{116}$$

which implies

$$\begin{aligned}&\lim_{|\mathbf{t}|\to 0} \text{var}(\varphi_{\widetilde{m}_g}(\mathbf{t}))\\ &\quad = M \lim_{|\mathbf{t}|\to 0}(1 - |\mathbb{E}[\exp(\mathrm{i}\boldsymbol{\epsilon}^T \mathbf{t})]|^2) \lim_{|\mathbf{t}|\to 0} \exp(-\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}) && (117)\\ &\quad = M \cdot 0 \cdot 1 = 0. && (118)\end{aligned}$$

■

A.3.5 Proof of Proposition 5

Proof of Proposition 5 : Convolving the discrete dot map with a uniform isotropic Gaussian kernel will get the density map

$$m = \sum_{j=1}^{T} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}), \tag{119}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}. \tag{120}$$

Then by Property 2, we have

$$\varphi_m(\mathbf{t}) = \sum_{j=1}^{T} \varphi_{\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})}(\mathbf{t}) = \sum_{j=1}^{T} \exp\left\{\mathrm{i}\langle \boldsymbol{\mu}_j, \mathbf{t}\rangle - \frac{\sigma^2 \langle \mathbf{t}, \mathbf{t}\rangle}{2}\right\}. \tag{121}$$

Let $D(a) = [-a, a]^2 - B(0, r)$ is a region in $\mathbb{R}^2$, then according to the definition of $\tilde{m}$, for non-trivial box A with zero-measure boundary, we have

$$\begin{aligned}
|m(A) - \tilde{m}(A)| &= \left| \lim_{a\to+\infty} \int_{[-a,a]^2} \int_A \frac{1}{(2\pi)^2} e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \varphi_m(\mathbf{t})\, d\mathbf{x} d\mathbf{t} - \lim_{a\to+\infty} \int_{[-a,a]^2} \int_A \frac{1}{(2\pi)^2} e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \varphi_{\tilde{m}}(\mathbf{t})\, d\mathbf{x} d\mathbf{t} \right| && (122)\\
&= \left| \lim_{a\to+\infty} \int_{D(a)} \int_A \frac{1}{(2\pi)^2} e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \varphi_m(\mathbf{t})\, d\mathbf{x} d\mathbf{t} \right| && (123)\\
&= \left| \lim_{a\to+\infty} \int_A \frac{1}{(2\pi)^2} \int_{D(a)} e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \varphi_m(\mathbf{t})\, d\mathbf{t} d\mathbf{x} \right| && (124)\\
&= \left| \lim_{a\to+\infty} \int_A \frac{1}{(2\pi)^2} \int_{D(a)} e^{-\mathrm{i}\langle \mathbf{t},\mathbf{x}\rangle} \left[\sum_{j=1}^{T} \exp\{\mathrm{i}\langle \boldsymbol{\mu}_j, \mathbf{t}\rangle - \frac{\sigma^2\langle \mathbf{t}, \mathbf{t}\rangle}{2}\} \right] d\mathbf{t} d\mathbf{x} \right| && (125)\\
&\le \lim_{a\to+\infty} \int_A \frac{1}{(2\pi)^2} \int_{D(a)} \exp\{-\frac{\sigma^2\langle \mathbf{t}, \mathbf{t}\rangle}{2}\} \left| \sum_{j=1}^{T} \exp\{\mathrm{i}\langle \boldsymbol{\mu}_j - \mathbf{x}, \mathbf{t}\rangle\} \right| d\mathbf{t} d\mathbf{x} && (126)\\
&\le \lim_{a\to+\infty} \int_A \frac{1}{(2\pi)^2} \int_{D(a)} \exp\{-\frac{\sigma^2\langle \mathbf{t}, \mathbf{t}\rangle}{2}\} \left[\sum_{j=1}^{T} \left|\exp\{\mathrm{i}\langle \boldsymbol{\mu}_j - \mathbf{x}, \mathbf{t}\rangle\}\right| \right] d\mathbf{t} d\mathbf{x} && (127)\\
&= \lim_{a\to+\infty} \int_A \frac{1}{(2\pi)^2} \int_{D(a)} \exp\{-\frac{\sigma^2\langle \mathbf{t}, \mathbf{t}\rangle}{2}\} * T\, d\mathbf{t} d\mathbf{x} && (128)\\
&= \frac{T\mathcal{L}(A)}{(2\pi)^2} \lim_{a\to+\infty} \int_{D(a)} \exp\{-\frac{\sigma^2\langle \mathbf{t}, \mathbf{t}\rangle}{2}\}\, d\mathbf{t} && (129)\\
&= \frac{T\mathcal{L}(A)}{(2\pi)^2} \int_{\mathbb{R}^2 - B(0,r)} \exp\{-\frac{\sigma^2\langle \mathbf{t}, \mathbf{t}\rangle}{2}\}\, d\mathbf{t} && (130)\\
&= \frac{T\mathcal{L}(A)}{(2\pi)^2} * \frac{2\pi * \exp\{\frac{-\sigma^2 r^2}{2}\}}{\sigma^2} && (131)\\
&= \frac{T\mathcal{L}(A) \exp\{\frac{-\sigma^2 r^2}{2}\}}{2\pi\sigma^2}. && (132)
\end{aligned}$$

(122) holds by Property 2; (124) holds by Fubini theorem; (126) holds by the absolute integral inequality; (128) holds by the truth $\left|\exp\{\mathrm{i}\langle \boldsymbol{\mu}_j - \mathbf{x}, \mathbf{t}\rangle\}\right| = 1$; (130) holds by monotone convergence theorem; (131) holds by calculating the integral in the polar system.

Finally, rearrange the term, we have

$$\frac{|m(A) - \tilde{m}(A)|}{\mathcal{L}(A)} \leq \frac{T\exp\{\frac{-\sigma^2 r^2}{2}\}}{2\pi\sigma^2}. \tag{133}$$

■

A.3.6 Proof of Proposition 6

Proof of Proposition 6 : By Property 3 (Inversion Formula), we have

$$\begin{aligned}
\tilde{m}(A) =&(2\pi)^{-2}\int_{[-a,a]^2}\int_A \varphi_m(\mathbf{t})\exp(-\mathrm{i}\langle\mathbf{t},\mathbf{x}\rangle)d\mathbf{x}d\mathbf{t} && (134)\\
=&(2\pi)^{-2}\int_A\int_{[-a,a]^2}\Sigma_{\{k=1,2,\cdots,N\}}m(\mathbf{p}^{(k)})\exp[\mathrm{i}\langle\mathbf{t},\mathbf{p}^{(k)}-\mathbf{x}\rangle]d\mathbf{t}d\mathbf{x} && (135)\\
=&(2\pi)^{-2}\int_A\Sigma_{\{k=1,2,\cdots,N\}}\Big[m(\mathbf{p}^{(k)})\int_{-a}^{a}\exp\{\mathrm{i}t_1(\mathbf{p}_1^{(k)}-\mathbf{x}_1)\}dt_1\int_{-a}^{a}\exp\{\mathrm{i}t_2(\mathbf{p}_2^{(k)}-\mathbf{x}_2)\}dt_2\Big]d\mathbf{x} && (136)\\
=&(2\pi)^{-2}\int_A\Sigma_{\{k=1,2,\cdots,N\}}\Big[m(\mathbf{p}^{(k)})[\frac{\exp\{(\mathbf{p}_1^{(k)}-\mathbf{x}_1)\mathrm{i}t_1\}}{(\mathbf{p}_1^{(k)}-\mathbf{x}_1)\mathrm{i}}\Big|_{-a}^{a}]*[\frac{\exp\{(\mathbf{p}_2^{(k)}-\mathbf{x}_2)\mathrm{i}t_2\}}{(\mathbf{p}_2^{(k)}-\mathbf{x}_2)\mathrm{i}}\Big|_{-a}^{a}]\Big]d\mathbf{x} && (137)\\
=&(2\pi)^{-2}\int_A\Sigma_{\{k=1,2,\cdots,N\}}\Big[m(\mathbf{p}^{(k)})\frac{2\sin\{(\mathbf{p}_1^{(k)}-\mathbf{x}_1)a\}}{\mathbf{p}_1^{(k)}-\mathbf{x}_1}*\frac{2\sin\{(\mathbf{p}_2^{(k)}-\mathbf{x}_2)a\}}{\mathbf{p}_2^{(k)}-\mathbf{x}_2}\Big]d\mathbf{x} && (138)\\
=&\sum_{k=1}^{N}m(\mathbf{p}_k)\int_A\prod_{d=1}^{2}\frac{\sin((\mathbf{p}_d^{(k)}-\mathbf{x}_d)a)}{\pi(\mathbf{p}_d^{(k)}-\mathbf{x}_d)}d\mathbf{x}. && (139)
\end{aligned}$$

(135) holds by Fubini theorem, and other equalities are regular integral calculations.

■

B. Derivation of Derivatives

In this section, we derive the derivatives of three losses regarding the output value $P(\mathbf{x})$ in subsection IV-B. Since $P(\mathbf{x})$ is real-value, we notice that

$$\begin{aligned}
\Re\{\Delta(t_1,t_2)\}^2 &= (\Re\{\varphi_{m_g}(t_1,t_2)\} - \Re\{\varphi_{m_p}(t_1,t_2)\})^2, && (140)\\
\Im\{\Delta(t_1,t_2)\}^2 &= (\Im\{\varphi_{m_g}(t_1,t_2)\} - \Im\{\varphi_{m_p}(t_1,t_2)\})^2, && (141)\\
\frac{\partial\Re\{\varphi_{m_p}(t_1,t_2)\}}{\partial P(\mathbf{x})} &= \frac{\partial\Re\{\sum_{\mathbf{y}\neq\mathbf{x}}\exp(\mathrm{i}\mathbf{y}^T\mathbf{t})P(\mathbf{y})\} + \Re\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})P(\mathbf{x})\}}{\partial P(\mathbf{x})} = \Re\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})\}, && (142)\\
\frac{\partial\Im\{\varphi_{m_p}(t_1,t_2)\}}{\partial P(\mathbf{x})} &= \frac{\partial\Im\{\sum_{\mathbf{y}\neq\mathbf{x}}\exp(\mathrm{i}\mathbf{y}^T\mathbf{t})P(\mathbf{y})\} + \Im\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})P(\mathbf{x})\}}{\partial P(\mathbf{x})} = \Im\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})\}. && (143)
\end{aligned}$$

Then by the chain rule of taking derivatives, since there is no item related to $P(\mathbf{x})$ in $\varphi_{m_g}(t_1,t_2)$, we have

$$\begin{aligned}
\frac{\partial\Re\{\Delta(t_1,t_2)\}^2}{\partial P(\mathbf{x})} &= \frac{\partial\Re\{\Delta(t_1,t_2)\}^2}{\partial\Re\{\varphi_{m_p}(t_1,t_2)\}} * \frac{\partial\Re\{\varphi_{m_p}(t_1,t_2)\}}{\partial P(\mathbf{x})} = -2\Re\{\Delta(t_1,t_2)\} * \Re\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})\}, && (144)\\
\frac{\partial\Im\{\Delta(t_1,t_2)\}^2}{\partial P(\mathbf{x})} &= \frac{\partial\Im\{\Delta(t_1,t_2)\}^2}{\partial\Im\{\varphi_{m_p}(t_1,t_2)\}} * \frac{\partial\Im\{\varphi_{m_p}(t_1,t_2)\}}{\partial P(\mathbf{x})} = -2\Im\{\Delta(t_1,t_2)\} * \Im\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})\}. && (145)
\end{aligned}$$

Then by the additivity rule of taking derivatives, we have

$$\frac{\partial \Re\{\Delta(t_1,t_2)\}^2 + \Im\{\Delta(t_1,t_2)\}^2}{\partial P(\mathbf{x})} \tag{146}$$

$$= -2\Re\{\Delta(t_1,t_2)\} * \Re\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})\} - 2\Im\{\Delta(t_1,t_2)\} * \Im\{\exp(\mathrm{i}\mathbf{x}^T\mathbf{t})\} \tag{147}$$

$$= 2\langle \mathbf{d}(\mathbf{t}), -f_{\mathbf{x}}(\mathbf{t})\rangle. \tag{148}$$

Let

$$h_{\mathbf{t}}(P(\mathbf{x})) = h_{(t_1,t_2)}(P(\mathbf{x})) = \Re\{\Delta(t_1,t_2)\}^2 + \Im\{\Delta(t_1,t_2)\}^2, \tag{149}$$

then finally, by the chain rule and additivity rule of taking derivatives, we have

$$\begin{aligned} \frac{1}{c^2}\frac{\partial \hat{l}_{\text{gchf}}(m_g,m_p)}{\partial P(\mathbf{x})} &= \sum_{\mathbf{t}\in\mathcal{R}} \frac{\partial\sqrt{h_{\mathbf{t}}(P(\mathbf{x}))}}{\partial h_{\mathbf{t}}(P(\mathbf{x}))} * \frac{\partial h_{\mathbf{t}}(P(\mathbf{x}))}{\partial P(\mathbf{x})} \\ &= \sum_{\mathbf{t}\in\mathcal{R}} \frac{1}{2\sqrt{h_{\mathbf{t}}(P(\mathbf{x}))}} * 2\langle \mathbf{d}(\mathbf{t}), -f_{\mathbf{x}}(\mathbf{t})\rangle \\ &= \sum_{\mathbf{t}\in\mathcal{R}} \frac{\langle \mathbf{d}(\mathbf{t}), -f_{\mathbf{x}}(\mathbf{t})\rangle}{\|\mathbf{d}(\mathbf{t})\|_2}, \end{aligned} \tag{150}$$

$$\begin{aligned} \frac{1}{c^2}\frac{\partial \bar{l}_{\text{gchf}}(m_g,m_p)}{\partial P(\mathbf{x})} &= \sum_{t_1\in\mathcal{R}_1} \frac{\partial\sqrt{\sum_{t\in\mathcal{R}_2} h_{(t_1,t)}(P(\mathbf{x}))}}{\partial \sum_{t\in\mathcal{R}_2} h_{(t_1,t)}(P(\mathbf{x}))} * \frac{\partial \sum_{t\in\mathcal{R}_2} h_{(t_1,t)}(P(\mathbf{x}))}{\partial P(\mathbf{x})} \\ &= \sum_{t_1\in\mathcal{R}_1} \Big[\frac{1}{2\sqrt{\sum_{t\in\mathcal{R}_2} h_{(t_1,t)}(P(\mathbf{x}))}} * 2\sum_{t\in\mathcal{R}_2}\langle \mathbf{d}(t_1,t), -f_{\mathbf{x}}(t_1,t)\rangle\Big] \\ &= \sum_{\mathbf{t}\in\mathcal{R}} \frac{\langle \mathbf{d}(\mathbf{t}), -f_{\mathbf{x}}(\mathbf{t})\rangle}{\mathcal{Q}(\mathbf{t})}, \end{aligned} \tag{151}$$

$$\begin{aligned} \frac{1}{c}\frac{\partial \tilde{l}_{\text{gchf}}(m_g,m_p)}{\partial P(\mathbf{x})} &= \frac{\partial\sqrt{\sum_{\mathbf{t}\in\mathcal{R}} h_{\mathbf{t}}(P(\mathbf{x}))}}{\partial \sum_{\mathbf{t}\in\mathcal{R}} h_{\mathbf{t}}(P(\mathbf{x}))} * \frac{\partial \sum_{\mathbf{t}\in\mathcal{R}} h_{\mathbf{t}}(P(\mathbf{x}))}{\partial P(\mathbf{x})} \\ &= \frac{1}{2\sqrt{\sum_{\mathbf{t}\in\mathcal{R}} h_{\mathbf{t}}(P(\mathbf{x}))}} * 2\sum_{\mathbf{t}\in\mathcal{R}}\langle \mathbf{d}(\mathbf{t}), -f_{\mathbf{x}}(\mathbf{t})\rangle \\ &= \sum_{\mathbf{t}\in\mathcal{R}} \frac{\langle \mathbf{d}(\mathbf{t}), -f_{\mathbf{x}}(\mathbf{t})\rangle}{\frac{1}{c}\tilde{l}_{\text{gchf}}(m_g,m_p)}. \end{aligned} \tag{152}$$

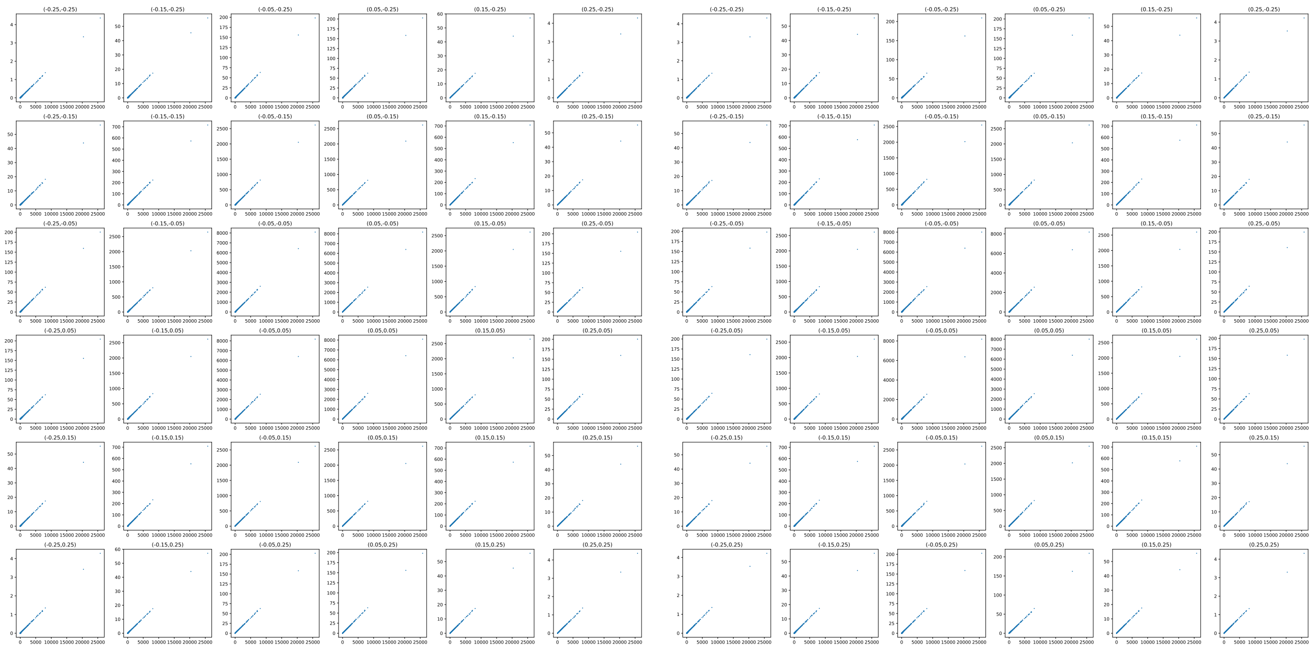
C. More Evidence of the Linearly Scaling Property of the Annotation Noise in the Frequency Domain

This section supplements more evidence about the linear scaling property of the annotation noise in the frequency domain. See Fig. 3.

D. Visualization of Density Maps and Localization Points

In this section, we visualize some generated density maps for crowd counting and localization as well as the localization results. They are shown in Fig. 4 and Fig. 5

The density maps generated by GCFL-CC are smoother and more robust to the local annotation noise, while the density maps generated by GCFL-CL are sharper and offer more local details for localization. Furthermore, the density maps for counting focus more on the area-level information, i.e., they cover all regions with people heads and show which regions



(a) (b)

Figure 3. Scatter plots of variance of the characteristic function versus people counts for different positions in the frequency domain for (a) the real part and (b) imaginary part. The position in the frequency domain are indicated in the title of each scatter plot. The variance scales linearly with the people counts, which validates the theoretical results.

are denser and which are sparser, while the density maps for localization carry more pixel-level information for giving the position of each individual. Thus the two types of density maps are mutually complementary. For crowd localization, our method can localize both the tiny heads as well as the big heads in both dense and sparse crowds. However, some extra predictions may be wrongly produced over non-heads.

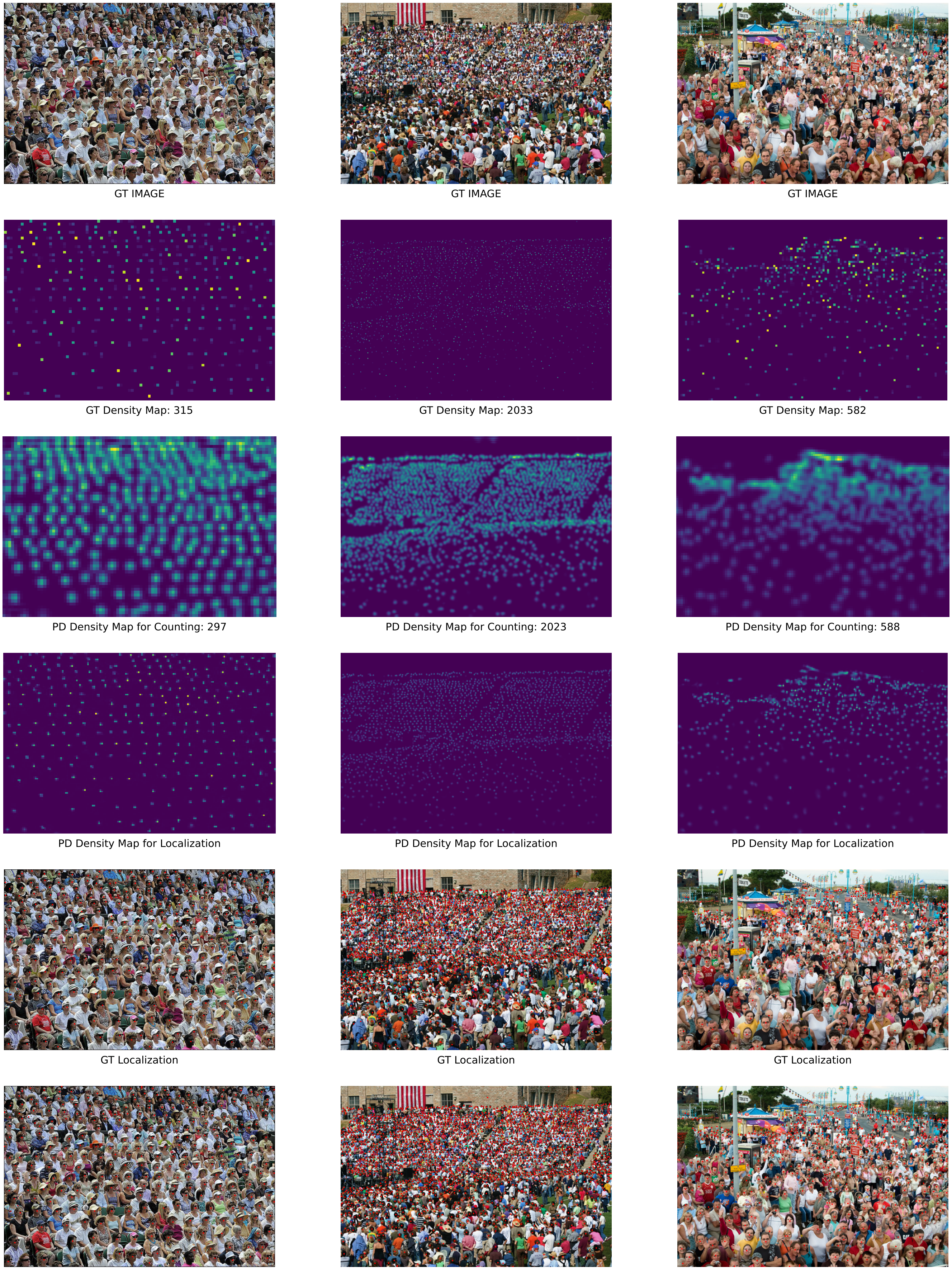


Figure 4. Results Visualization. The density maps generated by GCFL-CC are used for counting, while the density maps generated by GCFL-CL are used for localization. The former is smoother and more robust to the local annotation noise, while the latter is sharper and offers more local details for localization. GT means the ground truth, and PD means the prediction.

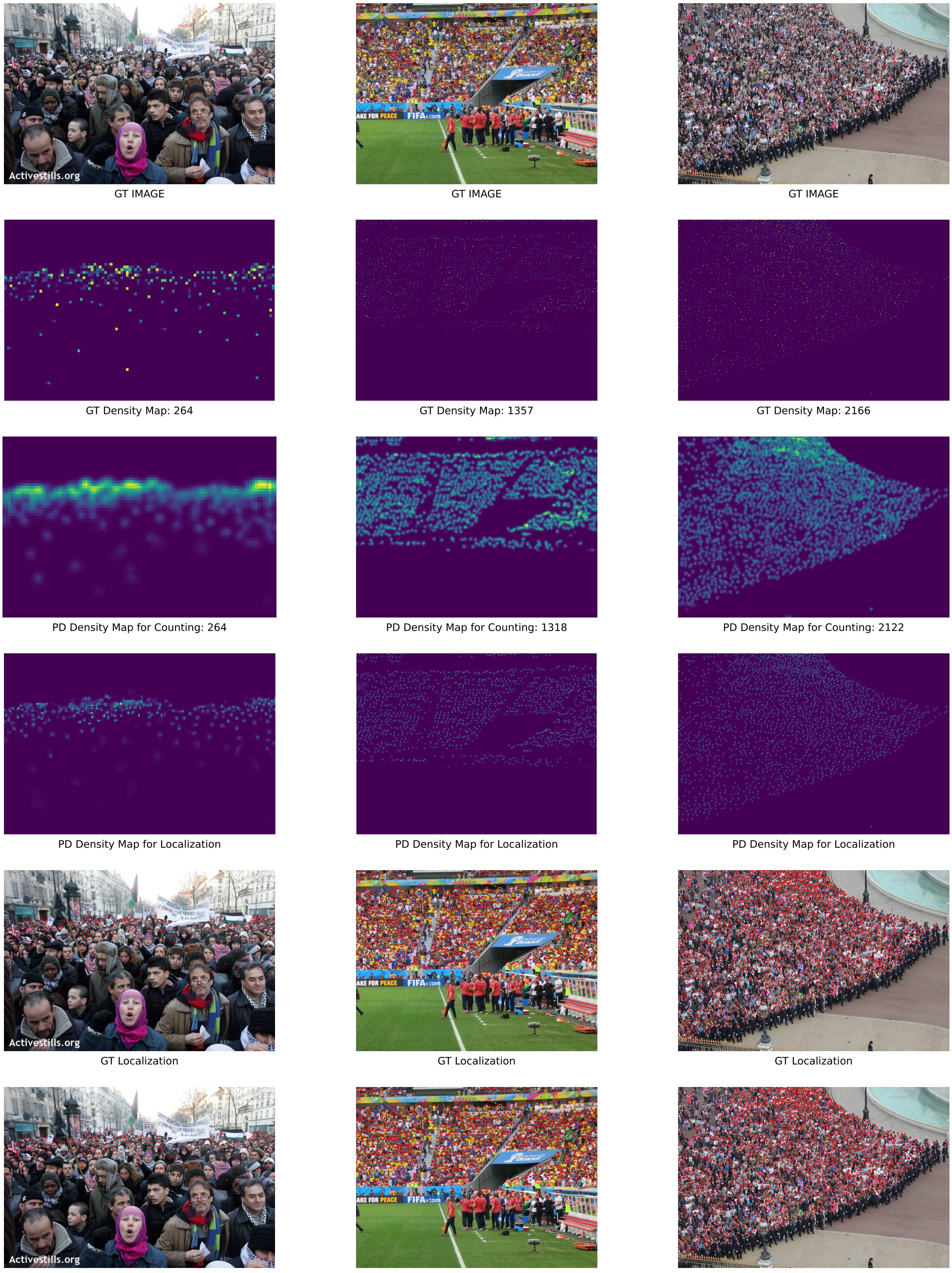


Figure 5. Results Visualization. The density maps generated by GCFL-CC are used for counting, while the density maps generated by GCFL-CL are used for localization. The former is smoother and more robust to the local annotation noise, while the latter is sharper and offers more local details for localization. GT means the ground truth, and PD means the prediction.