Rethinking Boundary Discontinuity Problem for Oriented Object Detection

Supplementary Material

7. Another understanding of ACM

7.1. From Complex Function to Polar Mapping

For the complex-exponential-based encoding proposed in the main text, leveraging Euler's formula allows for its transformation into a polar-coordinate mapping, as following:

$$\begin{aligned} z &= f(\theta) = e^{j\omega\theta} \\ &= \cos(\omega\theta) + j\sin(\omega\theta) \end{aligned} \tag{12}$$

$$\begin{aligned} \theta &= f^{-1}(z) = -\frac{j}{\omega}\ln z \\ &= \frac{1}{\omega}((\arctan2(z_{im}, z_{re}) + 2\pi) \bmod 2\pi) \end{aligned} \tag{13}$$

where $\omega \in \mathbb{R}^+$ is angular frequency, $\arctan2$ is another version of $\arctan$ with quadrant assignment, and z_{re}, z_{im} are real-part and imagine-part of complex coding $z \in \mathbb{Z}$, respectively. By hiding the complex mark of the encoding, we can regard it as a 2D polar coordinate encoding, as following:

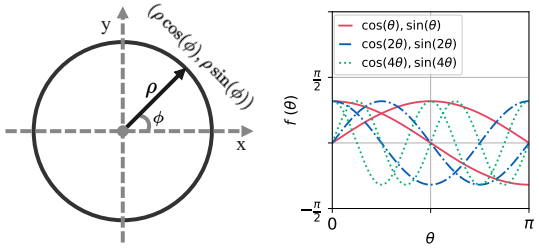
$$(z_x, z_y) = f(\theta) = (\cos(\omega\theta), \sin(\omega\theta)) \tag{14}$$

$$\begin{aligned} \theta &= f^{-1}(z_x, z_y) \\ &= \frac{1}{\omega}((\arctan2(z_y, z_x) + 2\pi) \bmod 2\pi) \end{aligned} \tag{15}$$

where $\omega \in \mathbb{R}^+$ is still angular frequency, $\arctan2$ is another version of $\arctan$ with quadrant assignment, z_x, z_y are x-axis-component and y-axis-component of 2D vector $\boldsymbol{z} \in \mathbb{R}^2$, respectively. This form is similar to continuous PSC encoding [41], but note that PSC cannot perform the above conversion.

7.2. Mathematical Meaning of Polar Mapping

In this perspective, encoding corresponds to the Cartesian coordinates of a unit vector, while decoding corresponds to the polar coordinates representation of the same unit vector. As is shown in Fig. 6a, given a vector with polar angle ϕ and radius(length) ρ in 2-dimensional space, it can be decompose as $(\rho\cos(\phi), \rho\sin(\phi))$ in Cartesian coordinates. When the radius ρ is fixed and ϕ is just considered in single period, the polar angle and Cartesian coordinates are one-to-one correspondences. Therefore, even leaving aside we can utilize this relationship to design f and obtain the corresponding f^{-1} as & Fig. 6b. In contrast, PSC coding [41] does not have such a clear mathematical meaning, so it needs to be experimentally determined to encoding length hyperparameters.



(a) (b)

Figure 6. Based on (a) polar coordinate decomposition, we define (b) a 2-dimensional wrapping function $f(\theta)$.

Table 6. Ablation study of different encoding mode.

$\omega=1$	$\omega=2$	$\omega=4$	HRSC2016		DOTA	
			AP_{50}	AP_{75}	AP_{50}	AP_{75}
			88.26	62.95	71.97	26.11
✓			90.44 (+2.18)	78.90 (+15.95)	73.51 (+1.54)	39.29 (+13.18)
	✓		**90.58** (+2.32)	86.12 (+23.17)	73.08 (+1.11)	39.62 (+13.51)
		✓	24.90 (-63.36)	20.82 (-42.13)	35.50 (-36.47)	17.29 (-8.82)
	✓	✓	90.55 (+2.29)	**87.77** (+24.82)	**74.51** (+2.54)	**40.49** (+14.38)

8. Determination of Angular Frequency

8.1. Perspective 1: Complex Function

To determine the appropriate ω, we discuss the relationship of $f_{box} \sim f_{obj}$ as following:

$$\begin{aligned} f_{box} &= e^{i\omega\theta_{box}} = e^{i\omega(\theta_{obj} \bmod \pi)} \\ &= \begin{cases} e^{i\omega\theta_{obj}}, & \theta_{obj} \in [0, \pi) \\ e^{i\omega\theta_{obj}} \cdot e^{-i\omega\pi}, & \theta_{obj} \in [\pi, 2\pi) \end{cases} \end{aligned} \tag{16}$$

1) When $\omega = 2$, $e^{-i\omega\pi} = 1$, then

$$\begin{aligned} f_{box} &= e^{i\omega\theta_{box}} \\ &= \begin{cases} e^{i\omega\theta_{obj}}, & \theta_{obj} \in [0, \pi) \\ e^{i\omega\theta_{obj}}, & \theta_{obj} \in [\pi, 2\pi) \end{cases} \\ &= f_{obj} \end{aligned} \tag{17}$$

2) When $\omega = 1$, $e^{-i\omega\pi} = -1$, then

$$\begin{aligned} f_{box} &= e^{i\omega\theta_{box}} \\ &= \begin{cases} e^{i\omega\theta_{obj}}, & \theta_{obj} \in [0, \pi) \\ -\ e^{i\omega\theta_{obj}}, & \theta_{obj} \in [\pi, 2\pi) \end{cases} \\ &= \begin{cases} f_{obj}, & \theta_{obj} \in [0, \pi) \\ -\ f_{obj}, & \theta_{obj} \in [\pi, 2\pi) \end{cases} \\ &= f_{obj} \cdot \text{sign}(\pi - \theta_{obj}) \end{aligned} \tag{18}$$

Through further derivation of the formula, we can find that **1)** When $\omega = 2$, Eq. (16) can be simplified to a straightforward $f_{box} = f_{obj}$. f becomes a consistent attribute for

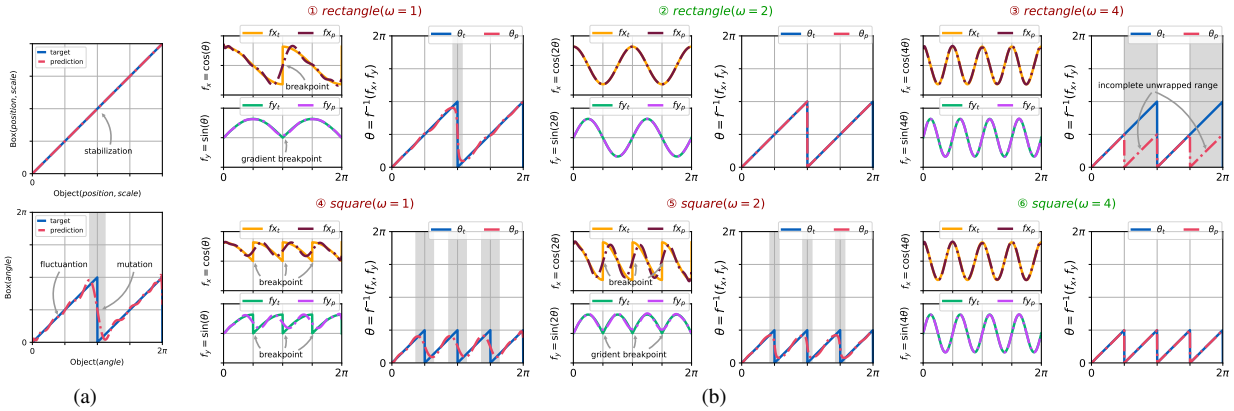


Figure 7. Waveform analysis: (a) original angular relationship between box and object. (b) based on polar-coordinate mapping, we define a 2-dimensional en/decoding function $f(\cdot)$. It will compound onto the original sawtooth wave $\theta_{box} = \theta_{obj} \pmod{\pi}$, and exhibits different effect for rectangular(top, $T = \pi$) & square-like(bottom, $T = \frac{\pi}{2}$) objects when angular frequency $\omega = 1, 2, 4$. The target and prediction are marked as solid line and dash line, respectively. Areas of inaccurate angular prediction are highlighted in gray. The optimal angular frequency for rectangular & square-like objects is 2 and 4, respectively.

both box and object, and it is perfectly in line with our design goals; **2)** When $\omega = 1$, Eq. (16) can be just simplified to a $f_{obj} \cdot \text{sign}(\pi - \theta_{obj})$. f_{box} and f_{obj} has a simple relationship but still with breakpoints; **3)** When $\omega \neq 2$ and $\omega \neq 1$, $e^{-i\omega\pi}$ is no longer a real factor, which makes Eq. (16) difficult to simplify, and $f_{box} \sim f_{obj}$ difficult to analyze. To sum up, we finally choose $\omega = 2$ in ACM.

8.2. Perspective 2: Polar Mapping

From the perspective of polar-coordinate mapping, ACM has more clear mathematical meaning and simple real number expression, so that we can carry out more direct analysis. Although the single dimensional $\cos(\omega\theta)$ and $\sin(\omega\theta)$ are many-to-one, integration of them can achieve a one-to-one effect in a higher dimension, making f an reversible transformation. Due to polar coordinate decoding can get unique angle only in a single cycle, $\omega\theta$'s range $[0, \omega\pi) \subseteq [0, 2\pi)$, so it is necessary to satisfy $\omega \leq 2$.

With encoding operation, original relationship $\theta_{box} \sim \theta_{obj}$ (Fig. 7a) becomes $f_{box} \sim \theta_{obj}$ (Fig. 7b), where f_{box} is the result of encoding function $f(\cdot)$ applied on θ_{box}. Thus, the waveform of $f_{box} \sim \theta_{obj}$ at $[0, 2\pi)$ is equivalent to repeating the encoded $\sin / \cos$ waveform at $[0, \pi)$ twice, due to the sawtooth wave of $\theta_{box} \sim \theta_{obj}$.

Obviously, the main issue of $\omega > 2$ (e.g. $\omega > 4$) lies in the incomplete decoding range, which will have a serious impact on angular prediction. In the valid angular frequency range, only when $\omega = 2$, both encoding components are smooth (continuous and with continuous gradient) at $\theta_{obj} = \pi$. It indeed completely eliminates the breakpoints in the components and thus completely solves the boundary problem; otherwise ($\omega \neq 2$), there is always be breakpoints in the components. Specially, when $\omega = 1$, although its cosinoidal component is continuous, its sinusoidal component do not include any breakpoints, which is equivalent to partially solving the problem. Therefore, decoded angle waveform is significantly closer to the perfect sawtooth wave compared with original prediction (Fig. 7a), but there is still a gap compared with $\omega = 2$.

By comparing prediction(dash lines) with ground-truth(solid lines) in Fig. 7b(top), we can find once ground-truth of wrapped value contains breakpoints, its prediction will become significantly worse, and according unwrapped angle, too. Finally, only $\omega = 2$ is the optimal choice that makes f continuous, differentiable, and reversible for rectangular objects.

8.3. Perspective 3: Experiments

Although we have analyzed from two different perspectives that $\omega = 2$ is the more appropriate angular frequency, the experiment is destined to be the more direct perspective. We conducted experiments with angular frequencies ($\omega = 1, 2, 4$), as is shown in Tab. 6. Compared with original KFIoU [38], enhanced version with ACM($\omega = 1$) get remarkable improvement since sinusoidal component in decomposition of the angle has no breakpoints for rectangles. It is consistent with the phenomenon (the smaller distortion area) observed in the en/decoding waveform diagrams. Moreover, ACM($\omega = 2$) eliminates all breakpoints in both two components in decomposition of the angle for rectangles, so it achieves greater improvement. Due to the inability to unwrap the full angular range for rectangles, ACM($\omega = 4$) exhibits severe performance degradation, especially for HRSC2016 dataset consisting entirely of large aspect ratio ships. When adopted the fusion of two angular frequencies simultaneously ($\omega = 2, 4$, details in Sec. 8.5

below), compared to $\omega = 2$, the results have little effect on large aspect ratios objects on HRSC2016 dataset and the results have slightly improved on DOTA dataset. This is because the DOTA dataset contains both large aspect ratio objects and square-like objects. Overall, AP (especially AP_{75}) can benefit a lot from ACM, which verifies our analysis. In following experiments, we adopt mixed angular frequencies ($\omega = 2, 4$).

8.4. Extension to Square-like Object

When the value of the object's width and height are close to each other, the bounding box will become a square-like from rectangle, which possesses stronger symmetry and leads to the period of B_{angle} shrinking from π to $\frac{\pi}{2}$. As a result, breakpoints will occur at more locations (i.e., $\frac{\pi}{2}$, π, and $\frac{3\pi}{2}$). In this case, if we continue to use Angle Correct Module proposed in the previous section, we should set ω to 4 accordingly, as is shown in Figure 7b(bottom). It is worth noting that when $\omega = 2$, breakpoints still exist in f_x at $O_{angle} = \frac{\pi}{2}, \pi, \frac{3\pi}{2}$, while f_y suffers from gradient breakpoints at these positions although it is continuous, which is similar to the case of $\omega = 1$ for the rectangle.

8.5. Generalization for Varied Aspect Ratio

Considering that the actual scene contains both square-like and rectangular objects, we attempt to use wrapped values with two frequencies (denoted as $f^{(\omega)}$, where $\omega = 2, 4$) simultaneously and fuse the unwrapped results to obtain a more accurate angular prediction. Similar strategies can also be found in previous work [41, 50]. For boxes rotated within $[0, \frac{\pi}{2})$, both $f^{(2)}$ and $f^{(4)}$ can unwrap correct angles. For boxes rotated within $[\frac{\pi}{2}, \pi)$, $f^{(2)}$ still unwraps correctly, while $f^{(4)}$'s unwrapped angles will be offset by one decoding-period $\frac{\pi}{2}$ to fall in $[0, \frac{\pi}{2})$. Therefore, ideally the difference between $\theta^{(2)} \in [0, \pi)$ and $\theta^{(4)} \in [0, \frac{\pi}{2})$ could only be 0 or $\frac{\pi}{2}$, but it only affects rectangle ($T = \pi$) and not square-like ($T = \frac{\pi}{2}$) in both training and inference phase. Note that $f^{(2)}$ suffers from breakpoints only for square-like rather than rectangle, and $f^{(4)}$ is immune to breakpoints for both rectangle and square-like, which just fails to correctly determine period range belonging to angles. Thus we can utilize coarse $\theta^{(2)}$ to correct the period range of fine $\theta^{(4)}$ as follows, where relaxation condition outside the parentheses are adopted in practice due to the independent errors of $f^{(2)}$ & $f^{(4)}$. Finally, we use this fusion strategy to adapt objects with varied aspect ratio.

$$\theta = \begin{cases} \theta^{(4)} + \frac{\pi}{2} & , if\ \theta^{(2)} - \theta^{(4)} > \frac{\pi}{4}\ (\theta^{(2)} - \theta^{(4)} = \frac{\pi}{2}) \\ \theta^{(4)} & , if\ \theta^{(2)} - \theta^{(4)} \leq \frac{\pi}{4}\ (\theta^{(2)} = \theta^{(4)}) \end{cases} \tag{19}$$

where the inequality condition (e.g. $\theta^{(2)} - \theta^{(4)} > \frac{\pi}{4}$) is just a relaxed version of the equality condition (e.g. $\theta^{(2)} - \theta^{(4)} = \frac{\pi}{2}$). The latter is just the judgment condition in the ideal state, while the former is actually adopted in practice, which brings better numerical stability.